HELSINGIN KAUPPAKORKEAKOULU
Liiketoiminnan teknologian laitos

JALKAPALLO-OTTELUN MAALIMÄÄRIEN JA LOPPUTULOKSEN TODENNÄKÖISYYKSIEN ENNUSTAMINEN VEDONLYÖNNISSÄ

Liikkeenjohdon systeemit
Pro Gradu –tutkielma
Johannes Rintaniemi 71366-9
Kevät 2005

Liiketoiminnan teknologian laitoksen laitosneuvoston kokouksessa 18 / 5 2005 hyväksytty arvosanalla erinomainen, 80p.

Merja Halme Tomi Seppälä
prof. Ph.D.

HELSINGIN KAUPPAKORKEAKOULU
Liikkeenjohdon systeemit
Pro gradu -tutkielma
Johannes Rintaniemi

TIIVISTELMÄ

16.5.2005

JALKAPALLO-OTTELUN MAALIMÄÄRIEN JA LOPPUTULOKSEN TODENNÄKÖISYYKSIEN ENNUSTAMINEN VEDONLYÖNNISSÄ

Tutkimuksen tavoitteet

Tutkimuksen teoreettisessa osassa luodaan katsaus vedonlyönnin peruskäsitteisiin ja alan aiempaan tutkimukseen sekä esitellään lyhyesti käytettyjen tilastollisten menetelmien teoriaa. Empiriaosassa muodostetaan ennustemalli, jonka avulla lasketaan 1X2-vedonlyönnin todennäköisyyksiä eri jalkapallosarjojen otteluihin. Tutkielman tavoitteena on selvittää onko pelkästään tilastolliseen analyysiin perustuvan ennustemallin avulla löydettyjen ylikertoimien avulla mahdollista kyetä voitolliseen tulokseen vedonlyöntimarkkinoilla.

Tutkimusmenetelmä

Lähtökohtana työssä on, että ottelun lopputuloksen todennäköisyydet määräytyvät pelaavien joukkueiden aiemmasta menestyksestä. Menestystä voidaan mitata useiden eri muuttujien avulla, joiden taustalla olevia piilorakenteita etsitään faktorianalyysin avulla. Näitä löydettyjä osatekijöitä mittaavien alkuperäisten muuttujien informaatiota tiivistetään pääkomponenttianalyysilla. Käyttämällä saatuja pääkomponenttipisteitä selittävinä muuttujina regressioanalyysissä, voidaan laskea todennäköisyysarviot yksittäisten otteluiden lopputuloksille. Vertaamalla laskettuja todennäköisyyksiä vedonvälittäjän tarjoamiin kertoimiin, voidaan markkinoilta etsiä kannattavia sijoituskohteita ja tutkia sijoitusten menestystä.

Lähdeaineisto

Tutkimuksessa käytetään työn tilastollisiin menetelmiin liittyvää koti- ja ulkomaista lähdekirjallisuutta. Lisäksi käytetään tieteellisiä artikkeleita ja kirjallisuutta urheiluvedonlyönnin todennäköisyyksien määrittämisestä. Empiriaosassa aineistona käytetään vedonvälittäjän tarjoamia pelikohteita kertoimineen eri maiden jalkapallosarjoista ajanjaksolta 07/04 – 02/05.

Tulokset

Tutkimus osoittaa, että jo muutamat osatekijät mittaavat kattavasti joukkueiden menestystä. Lisäksi havaitaan, että eri regressiomallien välillä on huomattavia eroja ennusteiden osumatarkkuudessa. Parhaan mallin osalta voidaan todeta, että pelkän tilastollisen analyysin avulla on mahdollista sijoittaa tuottavasti 1X2-vedonlyönnin pelikohteisiin.

Avainsanat

Urheiluvedonlyönti, faktorianalyysi, pääkomponenttianalyysi, regressioanalyysi.

SISÄLLYS

Työssä käytettyjä symboleita, merkintöjä ja lyhenteitä

$Y \sim D$	satunnaismuuttuja Y noudattaa todennäköisyysjakaumaa D
$\mathbf{A}^T$	matriisin tai vektorin A transpoosi
$\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$	diagonaalimatriisi, jossa arvot $\lambda_1, \lambda_2, \ldots, \lambda_p$ ovat diagonaalin alkioissa ja muiden alkioiden arvo 0
$\times$	karteesinen tulo
$\Gamma(.)$	Gamma-funktio
α	negatiivisen binomijakauman ylihajontaparametri
h	kotijoukkueen maalimäärä. Joukon suurin maalimäärä h_{max}, $H = \{0,1,\ldots,h_{max}\}$ ja $h \in H$.
a	vierasjoukkueen maalimäärä. Joukon suurin maalimäärä a_{max}, $A = \{0,1,\ldots,a_{max}\}$ ja $a \in A$.
M	maalimatriisi $\mathbf{M} = [m_{ha}]$, missä alkio m_{ha} ilmoittaa frekvenssin, että kotijoukkue tekee h maalia ja vierasjoukkue a maalia. Tällöin tapahtumaan kotivoitto (1) kuuluvat alkiot joille $h>a$, tapahtumaan tasapeli (X) alkiot joille $h=a$ ja tapahtumaan vierasvoitto (2) alkiot $h<a$.
P	todennäköisyysmatriisi $\mathbf{P} = [p_{ha}]$, missä p_{ha} ilmoittaa todennäköisyyden, että kotijoukkue tekee h maalia ja vierasjoukkue a maalia.
$\mathbf{C}_{\text{sarja } S}$	sarjalle S määritetty korjausmatriisi $\mathbf{C} = [c_{ha}]$
T	tulosmatriisi $\mathbf{T} = [t_{ha}]$, missä t_{ha} ilmoittaa todennäköisyyden, että kotijoukkue tekee h maalia ja vierasjoukkue a maalia.
NegBin	negatiivinen binomijakauma
BVDIP	bivariate diagonally inflated Poisson distribution (kahden muuttujan Poisson-yhteisjakauma, jossa diagonaalin alkioiden todennäköisyyksiä on kasvatettu)

1. Johdanto

Ihminen keksi uhkapelin jo esihistoriallisella ajalla. Argeologien löytämät todisteet uhkapelistä ovat entisen Babylonian ja Egyptin alueilta noin vuodelta 3600 eKr. Tuolloin pelaamiseen käytettiin astragalia, lähes nopan muotoista eläimen luuta. Satunnaisuuden, todennäköisyyden ja matematiikan yhteyttä pelaamiseen ei kuitenkaan ymmärretty vuosisatoihin. (Epstein 1967, 1)

Italialainen tiedemies Gerolamo Cardano käsitteli todennäköisyyteen liittyviä käsitteitä kirjassaan Liber de Ludo Alea. Hän pohti nopanheiton silmäluvun odotusarvoa, ja johti sen kokeellisesti kaikkien mahdollisten lopputulosten frekvenssien perusteella. Tämän jälkeen uhkapelistä tehtiin satunnaisesti erilaisia tutkimuksia ja kirjoituksia. Esimerkiksi Galileo teki 1600-luvun alussa tutkielman firenzeläisten aatelisherrojen pyynnöstä, jossa hän todisti matemaattisesti, miksi kolmen nopan heitossa yhteissumma kymmenen esiintyy useammin kuin yhdeksän. (mt. 2-3)

Varsinaisena todennäköisyyslaskennan syntymävuotena pidetään vuotta 1654. Tuolloin Antoine Gombauld esitti nopanheittoon liittyvän ongelman matemaatikko Blaise Pascalille. Tästä innostuneena Pascal kävi kirjeenvaihtoa aiheesta kolleganssa Pierre de Fermatin kanssa. Kirjeenvaihdon tulosten pohjalta Pascal laati periaatteet, joille nykyinen todennäköisyyslaskennan teoria perustuu. (mt. 3)

Nykyään vedonlyönti on globaali viihdeteollisuuden toimiala, joka kasvaa voimakkaasti. Myös asenne vedonlyöntiin on muuttunut. Enää sitä ei nähdä pelkästään uhkapelinä ja ajanvietteenä, vaan siihen suhtaudutaan yhä useammin eräänä mahdollisena sijoitusmuotona. Vedonlyöntimarkkinoiden ja rahoitusmarkkinoiden välillä on monia yhtäläisyyksiä. Ensinnäkin molemmilla markkinoilla sijoittajat, joilla on heterogeeniset tiedot ja odotukset, yrittävät hyötyä taloudellisesti käymällä kauppaa kohteiden tulevaisuuden arvoista epävarmuuden vallitessa. Toiseksi urheiluvedonlyönti ja rahoitusjohdannaisten kauppa ovat molemmat nollasummapeliä kahden sijoittajan välillä. Molemmilla markkinoilla liikkuu myös huomattavia pääomia. Vuonna 2002 neljän suurimman brittiläisen vedonvälittäjän liikevaihto oli lähes 10 miljardia puntaa. Maailman kenties johtavilla vedonlyöntimarkkinoilla Pohjois-Amerikassa vuotuisen liikevaihdon on arvioitu olevan vuosittain 380 miljardia dollaria (National Gambling Impact Study Commission, 1999).

Suuret pelivolyymit ovat yksi pääsyy siihen, että viime vuosina vedonlyönti ja sen sääntely eri maissa on ollut erityisen mielenkiinnon kohteena mediassa. Esimerkiksi Suomessa valtiollisen monopolin turvin pelitoimintaa järjestävä Veikkaus Oy ja ahvenanmaalainen PAF ovat tulkinneet arpajaislain sisältöä eri tavoin ja odottavat nyt asiaan oikeuslaitoksen päätöstä. Myös Euroopan Unionin alueella pelimonopoleja vastaan on hyökätty ja markkinoille syntyy jatkuvasti uusia vedonvälittäjiä. Markkinaselvityksessään Ernst & Young (2000) toteaa internetin, digitaali- ja mobiiliteknologian lisäävän vedonlyönnin suosiota entisestään. Digi-televisiossa rahapelit, urheilu ja muu viihde kytkeytyvät entistä tiiviimmin toisiinsa. Lähetysten interaktiivisuus tuo uusia mahdollisuuksia tuotekehitykselle erityisesti vedonlyöntipeleissä. Muutos tulee varmasti lisäämään alan maailmanlaajuista liikevaihtoa entisestään sekä mahdollistaa aivan uudentyyppisten vedonlyönnin pelimuotojen leviämisen massamarkkinoille.

Rahoitusmarkkinoilla toimivien sijoittajien pääasiallisena motiivina on maksimoida omistajien varallisuuden arvo halutun riskitason puitteissa. Sen sijaan valtaosalle vedonlyöjistä urheiluvedonlyönti tarjoaa mahdollisuuden kokea jännitystä pienellä riskillä. Tällöin vedonlyöjä on valmis häviämään satunnaisen pelipanoksensa, koska arvostaa sen aikaansaamaa lisäjännitystä esimerkiksi televisio-ottelun seuraamiselle enemmän kuin panoksen puhdasta taloudellista arvoa. Jatkuvasti kasvava joukko vedonlyöjiä pyrkii kohti ammattimaisempaa toimintaa, jossa vedonlyönti ei ole pelkästään ajanviete vaan kannattavaa toimintaa. Tällöin motiivina on maksimoida sijoitetun pääoman tuottoa, kuten missä tahansa sijoitustoiminnassa. Aivan kuten salkunhoitaja yrittää seuloa markkinoilta esiin aliarvostettuja sijoituskohteita, niin myös sijoitushakuinen vedonlyöjä yrittää löytää markkinoilta aliarvostettuja eli edullisia pelikohteita.

Vedonlyönnin pelikohteet voidaan jakaa kahteen ryhmään kertoimien muodostumisen perusteella: kiinteäkertoimiseen ja muuttuvakertoimiseen vedonlyöntiin. Kiinteäkertoimisessa vedonlyönnissä kertoimien laadinnasta vastaavat vedonvälittäjien kertoimenlaskijat. Heidän tarkoituksensa on yrittää määrittää pelikohteena olevan urheilutapahtuman eri tulosskenaarioiden todennäköisyydet ja asettaa kertoimet siten, että vedonvälittäjän voitto maksimoituu. Kattaakseen toiminnasta aiheutuneita kuluja ja pienentääkseen riskiä vedonvälittäjä tarjoaa kaikille tulosvaihtoehdoille hieman matalamman kertoimen kuin matemaattisen todennäköisyyden käänteisluku edellyttäisi. Kertoimien pienentäminen vähentää kuitenkin pelaajien mielenkiintoa ja peliliikevaihtoa, mikä tietyn pisteen jälkeen pienentää myös vedonvälittäjän kokonaistuottoja. Siksi vedonvälittäjän on tarkoin harkittava minkälaisella voittomarginaalilla hän määrittää kertoimensa (Vuoksenmaa 1999, 1).

Muuttuvakertoimisessa vedonlyönnissä vedonvälittäjä luo markkinapaikan, jolla kertoimet määräytyvät vapaasti. Tällöin vedonvälittäjä kerää reaaliajassa tietoa eri tulosvaihtoehdoille sijoitetuista panoksista, määrää niiden pohjalta eri tulosvaihtoehtojen kertoimet ja välittää ne edelleen pelaajien tietoon. Tällöinkin vedonvälittäjä pidättä osan panoksista itselleen provisiona markkinapaikan luomisesta. Tässä tapauksessa vedonvälittäjälle ei aiheudu riskiä kertoimien asetannasta tai pelin lopputuloksesta. Muuttuvakertoimisen vedonlyönnin tilanne vastaa pääomamarkkinoiden hinnanmuodostusta, jolloin eri sijoituskohteille määräytyy hinta markkinoilla sijoittajien yhteisen näkemyksen mukaisesti.

1.1 Tutkimuksen tavoitteet ja rajaukset

Vedonlyöntiä sijoitusmielessä harjoittava pelaaja yhdistää otteluiden analysoinnissa sekä tilastollista että kvalitatiivista analyysiä. Objektiivisen tutkimuksen kannalta tällainen asetelma on kuitenkin ongelmallinen, sillä kvalitatiivisen informaation osalta selkeiden ja yksikäsitteisten pelikriteereiden asettaminen on erittäin vaikeaa.

Tässä työssä tutkitaan jalkapallo-ottelun maalilukujen ennustamista pääkomponentti- ja regressioanalyysin avulla. Sovittamalla joukkueiden maaliluvuille sopiva todennäköisyysjakauma, voidaan määrittää todennäköisyydet ottelun eri lopputulemille kotivoitto (1), tasapeli (X) ja vierasvoitto (2). Tutkimuksen empiriaosassa näiden arvioiden laatua seurataan vertaamalla niitä vedonvälittäjän tarjoamiin 1X2-vedonlyönnin kertoimiin ja etsimällä kannattavia sijoituskohteita jalkapallosarjoissa. *Tutkielman tavoitteena on selvittää onko pelkästään tilastolliseen analyysiin perustuvan ennustemallin avulla löydettyjen ylikertoimien avulla mahdollista kyetä voitolliseen tulokseen vedonlyöntimarkkinoilla.*

Lähtökohtana työssä on, että ottelun lopputuloksen todennäköisyydet määräytyvät pelaavien joukkueiden aiemmasta menestyksestä. Joukkueiden menestystä mitataan useiden eri muuttujien avulla, joiden informaatiota tiivistetään pääkomponenttianalyysin avulla. Muodostettuja pääkomponentteja käytetään regressioanalyysissa selittävinä muuttujina maalilukujen odotusarvon ennustamisessa, joiden perusteella voidaan laskea ottelun 1X2-todennäköisyysarviot. Tutkielma on rajattu ennusteiden laskemiseksi 1X2-vedonlyöntiin ja empiriaosan tarkastelu rajataan koskemaan vain suurimpia jalkapallosarjoja sekä Ruotsin jääkiekkosarjaa, joissa toiminta on joukkueiden osalta ammattimaista.

1.2 Tutkielman rakenne

Johdanto-luvussa käydään läpi vedonlyönnin taustaa sekä tutkimuksen tavoitteet ja rakenne. Toisessa luvussa perehdytään urheiluvedonlyönnin peruskäsitteisiin ja aihepiirin aiempaan kirjallisuuteen ja tutkimukseen. Työssä käytettävien menetelmien teoriaa käsitellään luvussa kolme. Luvussa neljä esitellään työssä käytettävä aineisto ja rakennetaan empiriaosassa käytettävä ennustemalli. Luvussa viisi seurataan miten hyvin työssä esitetty malli menestyy käytännössä. Viimeisessä luvussa esitetään yhteenveto ja johtopäätökset työssä saaduista tuloksista.

2. Urheiluvedonlyönnin peruskäsitteitä ja aiempia tutkimuksia

Tässä luvussa esitellään urheiluvedonlyönnin yleisimmät pelimuodot. Lisäksi käydään läpi aiheeseen liittyviä peruskäsitteitä. Luvun lopuksi kirjallisuuskatsauksessa perehdytään aiheen aiempiin tutkimuksiin.

2.1 Yleisimmät pelimuodot ja esimerkkejä [1]

Vedonlyönnin suosion kasvaessa myös erilaiset pelimuodot ovat lisääntyneet. Tässä kappaleessa esitellään tämän hetken yleisimmät pelimuodot.

1X2-vedonlyönti

Pelimuodossa pelikohteena on ottelun lopputulos, joka voi olla kotivoitto (1), tasapeli (X) tai vierasvoitto (2). Lopputuloksella tarkoitetaan tässä varsinaisen peliajan tulosta, jolloin esimerkiksi jääkiekossa jatkoajalla tapahtunut ratkaisu on aina merkiltään X.

Tasoitusveto

Tämä pelimuoto on läheistä sukua 1X2-vedoille, koska pelikohteena on jälleen ottelun lopputulos. Tasoitus tarkoittaa, että toinen joukkueista saa tasoitusta maalimäärään. Yleisimmin tasoitus on sellainen, että lopputulos voi olla ainoastaan koti- tai vierasvoitto.

Esimerkki: Jalkapallo-ottelu Aston Villa- Manchester United, jossa Aston Villa saa +0.5 maalia tasoitusta eli Aston Villan saamaan maalimäärään lisätään lopuksi 0.5 maalia.

Jos ottelu päättyy tulokseen 0-0, on lopullinen tulos tasoituksen jälkeen 0.5-0, ja oikea merkki tasoitusvedossa on 1 eli kotivoitto.

Tulosveto

Tulosvedossa pelikohteena on ottelun lopullinen tulos eli yleensä maalimäärä. Tällöin pelaajan on voittaakseen tiedettävä molempien joukkueiden tekemien maalien määrä.

---

[1] Luvun tiedot perustuvat internet-sivustoihin http://www.bettingadvice.com ja http://www.annabet.com.

Voittajaveto

Pelimuodossa kohteena voi olla lähes mikä tahansa kilpailu. Pelaajan tarkoituksena on onnistua veikkaamaan lopullinen voittaja. Suomessa voittajavedot rajoittuvat lähinnä urheilukilpailuihin, mutta maailmalla on mahdollista veikata lähes minkä tahansa kilpailun voittajaa. Esimerkiksi australialainen Centrebet tarjosi mahdollisuutta veikata eduskuntavaalien 2003 suurinta puoluetta.

Money-line

Tämä pelimuoto on sukua 1X2- ja tasoitusvedoille. Tarkoituksena on veikata ottelun lopputulos. Ottelun päättyessä tasapeliin panokset palautetaan, jolloin mahdollisia tuloksia ovat vain kotivoitto (1) ja vierasvoitto (2).

Over/Under

Pelimuotoa voidaan pitää tulosvedon johdannaisena. Tarkoitus on veikata tehdäänkö ottelussa yhteensä yli vai alle x maalia. Raja on asetettu siten, että vain yli tai alle ovat mahdollisia lopputuloksia.

Esimerkki: Tehdäänkö ottelussa Aston Villa- Manchester United yli vai alle 2,5 maalia.

Head-to-Head

Pelin kohteena on kilpailijoiden keskinäinen paremmuus, josta tapahtumassa ei varsinaisesti kilpailla.

Esimerkki: Formula 1-osakilpailu, kumpi sijoittuu paremmin M.Schumacher vai K.Räikkönen.

Exotics

Pelikohteena voi olla mikä tahansa otteluun liittyvä tapahtuma, joka ei kuulu aikaisempien pelimuotojen joukkoon. Kohteet vaihtelevat vedonvälittäjän ja ottelun mukaan, joten kattava listaaminen on mahdotonta.

Esimerkkejä: avausmaalin tekoaika, ottelun tilanne puoliaika/lopputulos, koko sarjakierroksen maalien lukumäärä.

Indeksivedonlyönti

Nopeimmin kasvava vedonlyönnin muoto on indeksivedonlyönti, joka sai alkunsa 1980-luvun loppupuolella Englannissa. Suurin ero edellisiin pelimuotoihin nähden on, että indeksivedoissa sekä voiton että tappion määrä selviää vasta ottelun päätyttyä. Tästä syystä myös panokset ja voitot maksetaan vasta ottelun päätyttyä. (Jackson 1994, 309-310)

Esimerkki: Vedonvälittäjä arvioi koripallo-ottelussa syntyvän yhteensä 145 pistettä ja tarjoaa pelaajille indeksiä 140-150. Pelaaja A uskoo pisteitä syntyvän alle 140, joten hän myy indeksin osuuksia. Pelaaja B uskoo puolestaan pisteitä syntyvän yli 150 ja hän ostaa osuuksia indeksistä.

Panos on 1 euroa / osuus. Ottelu päättyy 82-80 eli pisteitä syntyy yhteensä 162.
Pelaaja A oli väärässä ja hän häviää: 162-140 = 22 osuutta * 1 € = 22 €.
Pelaaja B oli oikeassa ja hän voittaa: 162-150 = 12 osuutta * 1 € = 12 €.

Vedonvälittäjä saa provisiona summien erotuksen eli 10 €. Tässä indeksi määrää vedonvälittäjän voittomarginaalin eli mitä pienempi indeksi, sitä pienempi on provisio. Jos ottelu päättyy tulokseen, joka on indeksin välillä, niin ei kumpikaan pelaaja voita. Tällöin ainoastaan vedonvälittäjä saa oman provisionsa. Oletetaan, että esimerkkimme ottelu päättyy 75-70 eli pisteitä yhteensä 145. Tällöin pelaaja A on väärässä ja häviää 145-140 = 5 € ja pelaaja B häviää 150-145 = 5 €, jolloin vedonvälittäjä saa provisionsa 10 €.

2.2 Vedonlyönnin perusteita ja käsitteitä [2]

Vedonvälittäjä tarjoaa pelaajille kertoimia eli todennäköisyysarvioita erilaisista urheilutapahtumista, joita pelaaja vertaa omiin arvioihinsa. Mikäli pelaaja löytää mielestään väärin olevia arvioita, jotka ovat hänen näkökulmastaan edullisia hän ostaa kyseisen arvion vedonvälittäjältä eli sijoittaa rahansa kyseiselle vedolle.

Vedonlyönti on hyvin samankaltaista kuin pörssikauppa, jossa spekulaation kohteena ovat yritysten tulokset ja pohjana niiden osakehintoihin sisältyvät odotusarvot. Erityisesti muutaman viime vuoden aikana on esitelty laajasti osakeanalyytikkojen salkkujen menestystä verrattuna markkinaindekseihin tai jopa sattumanvaraisesti valittuihin osakesalkkuihin. Tällä tavoin on

[2] Luvun tiedot perustuvat käsitteiden osalta internet-sivustoihin http://www.bettingadvice.com ja http://www.annabet.com.

joissakin tapauksissa kyseenalaistettu sekä analyytikkojen osaaminen että koko analyysin mielekkyys. Vaikka analyytikoissa on varmasti eroja, niin laadukkaalla analyysillä pystyy erittäin todennäköisesti lyömään summittaisen salkun aikajänteen ollessa riittävän pitkä. Samalla tavalla myös urheiluvedonlyönnissä on mahdollista hyvän analyysin avulla parantaa menestymisen mahdollisuuksia.

Todennäköisyysarvio (p)

Tapahtuman ennustettu todennäköisyys.

Alustava kerroin (o)

Ennustetun todennäköisyyden käänteisluku.

Deduktio (d) & palautusprosentti (r)

Deduktio on osa, jonka vedonvälittäjä pidättää itselleen panoksista. Sillä vedonvälittäjät kattavat toiminnasta aiheutuneita kuluja ja hyödyntävät sitä riskinhallinnassa. Pelimuodosta ja vedonvälittäjästä riippuen deduktion suuruus on 5 - 50 %. [3]

Palautusprosentti on osuus, jonka vedonvälittäjä maksaa panoksista voittoina pelaajille (palautusprosentti = 100 % - deduktio). PAF:in deduktio 1X2-vedonlyönnissä on 10 %. Pitkällä aikavälillä PAF maksaa jokaisesta pelatusta eurosta keskimäärin 90 senttiä voittoina pelaajille ja pitää itsellään 10 senttiä.

Lopullinen kerroin

Lopullinen kerroin saadaan alustavan kertoimen ja palautusprosentin tulona tai vaihtoehtoisesti palautusprosentin ja todennäköisyysarvion osamääränä. Se ilmoittaa kuinka moninkertaisesti pelaaja saa panoksensa takaisin vedon mennessä oikein. Esimerkiksi kerroin 2.00 antaisi osuessaan 100 € panoksella takaisin 200 €. Pelaajan nettovoitto olisi tällöin 100 €.

Odotusarvo & ylikerroin

Odotusarvo ilmaisee vedon odotettavissa olevan tuoton pitkällä aikavälillä. Odotusarvo lasketaan kertoimen ja todennäköisyysarvion tulona. Pelaajan kannalta kohde on kannattava, jos sen odotusarvo on yli yksi. Tällaista kerrointa kutsutaan ylikertoimeksi.

---

[3] Vedonlyöntipörssien (esimerkiksi www.betfair.com) kohdalla deduktio on yleensä alle 5 %. Tämä johtuu siitä, että pörssit luovat muuttuvakertoimisen vedonlyönnin tapaan markkinapaikan eivätkä kanna riskiä ottelun lopputuloksesta.

Taulukossa 1 on esitetty kootusti eri käsitteiden suhteet käyttäen esimerkkinä kolikonheittoa.

Taulukko 1: Vedonlyönnin käsitteiden yhteenveto heitettäessä harhatonta kolikkoa.

	Klaava	Kruuna
Todennäköisyysarvio p	50 %	50 %
Alustava kerroin o = 1/p	2,00	2,00
Deduktio d	10 %	10 %
Palautusprosentti r = 100 % - d	90 %	90 %
Lopullinen kerroin k = r/p = o*r	1,80	1,80

1X2-vedonlyönninssä pelaaja veikkaa päättyykö ottelu kotivoittoon, tasapeliin vai vierasvoittoon. Esimerkki: Vedonvälittäjä on antanut kertoimet ottelun eri lopputuloksille

	(1)	(X)	(2)
Manchester United-Liverpool	1,65	3,40	4,65.

Käyttämällä hyväksi taulukon 1 kaavoja voidaan laskea vedonvälittäjän omat todennäköisyysarviot ottelun lopputuloksille. Lasketaan aluksi yhteen kertoimien käänteislukujen summat

$$\frac{1}{1{,}65}+\frac{1}{3{,}4}+\frac{1}{4{,}65}=1{,}1152\,.$$

Teoreettinen palautusprosentti saadaan edellä lasketun summan käänteislukuna eli tässä tapauksessa $\frac{1}{1{,}1152}=0{,}897$ jolloin deduktion osuus on 1-0,897 = 0,103. Aivan kuten taulukon 1 tapauksessa vedonvälittäjä maksaa pelaajille voittoina pitkällä aikavälillä keskimäärin 90 % panoksista.

Olettaen, että vedonvälittäjä on jakanut deduktion tasan kaikkien eri lopputulos vaihtoehtojen kesken[4], voidaan vedonvälittäjän omat todennäköisyysarviot p_i ottelun eri lopputuloksille i (i=1,X,2) laskea kaavalla

[4] Käytännössä tämä ei aina pidä paikkaansa. Jotta deduktio toimisi halutulla tavalla, tulisi panosten jakautua ottelun eri tulosvaihtoehdoille vedonvälittäjän asettamien todennäköisyysarvioiden suhteessa. Tällöin vedonvälittäjät saattavat huomioida joukkopsykologian vaikutusta panosten jakautumisessa ja jakaa deduktion epätasaisesti eri vaihtoehdoille. Esimerkiksi suomalaisten vedonlyöjien panostukset kasautuvat maaotteluissa Suomen joukkueen hyväksi.

$$p_i = \frac{r}{k_i} \tag{2.1}$$

missä r on vedonvälittäjän palautusprosentti ja k_i kerroin lopputulokselle i. Esimerkkimme tapauksessa todennäköisyysarvioiksi saadaan

	(1)	(X)	(2)
Manchester United-Liverpool	0,543	0,264	0,193.

2.3 Vedonlyönnin aiempia tutkimuksia

Vedonlyönnin suuren suosion seurauksena aiheesta on julkaistu runsaasti kirjoja. Suurin osa näistä on kuitenkin luokiteltavissa viihdekirjallisuudeksi, joissa aiheen käsittely ei ole tyyliltään tieteellistä. Vedonlyönnistä julkaistuissa tieteellisissä artikkeleissa aihetta on käsitelty yleensä vedonlyöntimarkkinoiden tehokkuuden, optimaalisen panostuksen tai todennäköisyyksien määrittämisen näkökulmasta. Tässä työssä tarkastelun ulkopuolelle on jätetty vedonlyöntimarkkinoiden tehokkuutta ja panostamista käsittelevä aineisto.

Vedonlyöntiin liittyy runsaasti satunnaisuutta eikä paraskaan malli kykene selittämään ottelun lopputulosta. Pelaajan on seurattava tarkkaan eri joukkueiden menestymistä eli tiedettävä miten usein joukkue on voittanut, pelannut tasan ja hävinnyt sekä kuinka paljon kukin joukkue on tehnyt ja päästänyt maaleja. Nämä tekijät on varsin helppo muuttaa numeeriseen muotoon. Sen sijaan todennäköisyyksiin vaikuttavat useat muutkin asiat, joiden matemaattinen formalisointi on melko vaikeaa. Tällaisia tekijöitä ovat esimerkiksi loukkaantumiset, pelikiellot, valmentajavaihdokset ja motivaatiotekijät. Tämän haasteellisuuden ovat huomanneet myös monet tutkijat. Ehkäpä siksi vedonvälittäjät hyödyntävät tilastollisia menetelmiä kerrointenlaskennassa yllättävän vähän, kuten Jackson (1994) artikkelissaan toteaa.

Usein tutkimustilanteita on yksinkertaistettu, keskitytty johonkin tiettyyn pelin osatekijään ja yritetty selvittää sen vaikutusta ottelun kulkuun ja lopputulokseen. Reep & Benjamin (1968) tutkivat syöttöjen tapaa ja määrää englantilaisessa jalkapallossa. Ridder et al. (1994) analysoivat ulosajojen vaikutusta ottelun lopputulokseen. Kuonen (1996) mallinsi pudotuspelejä, joissa voittaja jatkaa ja häviäjä tippuu pois kilpailusta. Barnett & Hilditch (1993) tutkivat keinonurmen vaikutusta ottelun lopputulokseen. Clarke & Norman (1995) tutkivat kotiedun voimakkuutta ja sen vaikutusta ottelun tulokseen. Dixon & Robinson (1998) tutkivat koti- ja vierasjoukkueen maalintekointensiteetin vaihtelua ottelun aikana. Koning et al. (2003) kehittivät yhdistetyn simulointi-tilastomallin, jolla he tutkivat maa- ja seurajoukkueiden turnausmenestystä. Dobson & Goddard (2003) tutkivat voitto- ja tappioputkien esiintyvyyttä Englannin Valioliigassa[5]. Hill (1974) tutki asiantuntijoiden ennen kautta tekemien ennusteiden ja lopullisten sarjataulukoiden välistä korrelaatiota. Valmentajanvaihdoksen merkitystä joukkueen tulevaan menestykseen ovat tutkineet Bruinshoofd & Weel (2003) Hollannin jalkapallon pääsarjassa ja Audas et al. (1997) Englannin Valioliigassa.

Paakkulainen (1995) tutki Pitkävetokerrointen osumatarkkuutta eri lajeissa. Hän selvitti, oliko Veikkaus Oy onnistunut kerrointen määrittelyssä paremmin jossain tietyissä lajiryhmissä ja oliko viikonpäivällä merkitystä kertoimen tarkkuuteen. Lee (1999) tutki negatiivisen binomijakauman soveltuvuutta rugbyotteluiden maalimäärien mallintamiseen.

Jalkapallo-ottelun lopputuloksen mallintamisesta on useita tutkimuksia. Osassa lopputulosta on ennustettu suoraan ja toisissa taas on keskitytty ennustamaan joukkueiden maalimääriä, joiden perusteella voidaan laskea ennusteet myös lopputuloksille. Tiettävästi ensimmäinen tutkimus on Moroneyn (1956) tekemä tilastollinen analyysi. Hän sovitti kirjassaan Poisson-jakaumaa jalkapallon maalimäärille mutta totesi, ettei yhteensopivuus jakaumaan ollut tilastollisesti merkitsevä. Hän esitti ratkaisuksi tilanteeseen Poisson-muunnoksen[6] käyttöä. Reep et al. (1971) tutkivat negatiivisen binomijakauman sovittamista sekä jalkapalloon että moniin muihin pallopeleihin. He totesivat jakauman soveltuvan hyvin jalkapallon, jääkiekon ja baseballin pistemäärien mallintamiseen.

Huolimatta Poisson-jakauman aiemmista heikoista tuloksista, tutki Maher (1982) Poisson-mallia lisää. Hän vertasi useiden kausien ajalta koti- ja vierasmaalien odotettuja ja havaittuja frekvenssejä ja totesi niiden noudattavan Poisson-jakaumaa. Lee (1997) rakensi Poisson-regressio mallin, jonka avulla hän simuloi pelikausia ja tutki ansaitsiko Manchester United mestaruuden kaudella 1996-1997. Hän tutki myös koti- ja vierasmaalien riippumattomuutta kokonaisen pelikauden tuloksilla ja totesi ne tilastollisesti riippumattomiksi. Myös Dixon & Coles (1997) rakensivat Poisson-regressio mallin, mutta heidän mallinsa sisälsi hyökkäys- ja puolustuskyvykkyyden lisäksi aikaparametrin. Koning (2000) tutki mallia, jossa selitettiin suoraan ottelun lopputulosta ordered probit -regression avulla.

---

[5] Ammattilaissarja Valioliiga on Englannin korkein sarjataso jalkapallossa.
[6] Moroneyn Poisson-muunnos oli itse asiassa negatiivinen binomijakauma (Reep et al. 1971, 624-625).

Rue & Salvesen (2000) huomioivat sarjatilanteen dynaamisen luonteen ja kehittivät Poisson-jakaumalle perustuvan mallin, jonka parametrisoinnissa he hyödynsivät Markovin ketjujen simulointiin perustuvaa Monte Carlo -menetelmää. Myös Crowder et al. (2002) esittivät samansuuntaisen mallin, jossa joukkueiden voimasuhteiden oletettiin muodostuvan stokastisen prosessin tuloksena.

Viimeisimpiä tutkimuksia ovat esittäneet Hirotsu & Wright (2003). He esittivät Markov-prosessiin pohjautuvan mallin, jossa maalimäärien lisäksi huomioitiin myös joukkueiden pallonhallinta. Mallin avulla he tutkivat joukkueiden menestystä kauden aikana, riippuvuutta kotiedusta ja menestystä tiettyjä joukkueita vastaan. Goddard & Asimakopoulos (2003) rakensivat ordered probit -regressiomallin, jonka avulla he etsivät kannattavia vedonlyöntikohteita englantilaisesta jalkapallosta. Tulosten mukaan valitsemalla positiivisen odotusarvon kohteita kausien lopuilta kyettiin voitolliseen vedonlyöntiin. Haas (2003) tutki Saksan Bundesliigan[7] joukkueiden tehokkuutta DEA:n avulla. Tuotoksia mitattiin kerättyjen pisteiden, liikevoiton ja stadionien käyttöasteen avulla. Tuloksien mukaan sarjasijoitus ja DEA:n mukainen tehokkuus eivät korreloineet tilastollisesti merkitsevästi.

Aihetta on käsitelty myös opinnäytetöissä. Raitanen (1999) tutki eri pallopelien tulosten tilastollista mallintamista käyttäen hyväksi yleistettyä Poisson-jakaumaa. Marttinen (2001) vertasi erilaisten ennustemallien toimivuutta jalkapallo-otteluiden maalimäärien ja tulosten ennustamisessa, keskittyen lähinnä erilaisten Poisson-regressio mallien vertailuun. Juva (2003) esitti Poisson-jakauman sijasta Markov-mallin lopputulosjakauman ennustamiseksi.

[7] Ammattilaissarja Bundesliiga on Saksan korkein sarjataso jalkapallossa.

3. Käytettyjen menetelmien esittely

Tässä kappaleessa käydään lyhyesti läpi työssä käytettyjä tilastollisia menetelmiä.

3.1 Pääkomponenttianalyysi [8]

Pääkomponenttianalyysin tehtävä on löytää muuttujien toisistaan riippumattomia lineaarisia yhdistelmiä, jotka keräävät mahdollisimman suuren osan alkuperäisten muuttujien kokonaisvaihtelusta. Teknisenä tavoitteena pidetään usein ilmiötä kuvaavien muuttujien lukumäärän vähentämistä.

Etsitään muuttujien $\mathbf{X}=(X_1,X_2,\dots,X_p)$ sellaista lineaarista yhdistelmää

$$Y = b_1X_1 + b_2X_2 + \cdots + b_pX_p = \mathbf{b}^T\mathbf{X}, \tag{3.1}$$

jonka varianssi on mahdollisimman suuri. Tehtävä ei ole mielekäs, ellei kertoimien $b_1,b_2,\dots,b_p$ absoluuttista suuruutta säädellä jollain tavoin. Teknisesti helpoin rajoitus on asettaa kertoimien neliösumma ykköseksi eli $\mathbf{b}^T\mathbf{b}=1$.

Olkoon $E(\mathbf{X})=\boldsymbol{\mu}$ ja $\mathrm{cov}(\mathbf{X})=\boldsymbol{\Sigma} \geq 0$. Tällöin tehtävänä on maksimoida

$$\mathrm{var}(Y) = \mathbf{b}^T\mathrm{cov}(\mathbf{X})\mathbf{b} = \mathbf{b}^T\boldsymbol{\Sigma}\mathbf{b} \text{ ehdolla } \mathbf{b}^T\mathbf{b} = 1.$$

Tulos saadaan matriisin $\boldsymbol{\Sigma}$ spektraalihajotelmasta,

$$\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Lambda}\mathbf{B}^T = \lambda_1\mathbf{b}^{(1)}\mathbf{b}^{(1)T} + \lambda_2\mathbf{b}^{(2)}\mathbf{b}^{(2)T} + \cdots + \lambda_p\mathbf{b}^{(p)}\mathbf{b}^{(p)T}, \tag{3.2}$$

missä $\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ ja $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Maksimointitehtävän ratkaisu saadaan kun $\mathbf{b}=\mathbf{b}^{(1)}$, tällöin neliömuoto $\mathbf{b}^T\boldsymbol{\Sigma}\mathbf{b}$ saavuttaa maksimiarvon λ_1 ja muuttujaa $Y_1= \mathbf{b}^{(1)T}\mathbf{X}$ sanotaan ensimmäiseksi pääkomponentiksi.

[8] Luku perustuu kokonaisuudessaan lähteeseen Mustonen (1995).

Tehtävää voidaan jatkaa etsimällä uutta yhdistettyä muuttujaa $Y=\mathbf{b}^T\mathbf{X}$, joka on korreloimaton ensimmäisen pääkomponentin Y_1 kanssa ja jonka varianssi ehdolla $\mathbf{b}^T\mathbf{b}=1$ on maksimaalinen. Tässä työssä käytetään vain ensimmäistä pääkomponenttia.

Pääkomponenttianalyysin tulos ilmaistaan tavallisimmin pääkomponenttimatriisina **F**. Kun analyysi tehdään korrelaatiomatriisista, on pääkomponenttimatriisin alkio f_{ij} yksinkertaisesti

$$\rho(X_i, Y_j) = f_{ij}$$

eli muuttujan X_i ja pääkomponentin Y_j välinen korrelaatiokerroin.

Usein pääkomponenttianalyysin tuloksia käytetään jatkotarkasteluissa ja ollaan kiinnostuneita havainnoille estimoiduista pääkomponenttien pistemääristä. Käytettäessä vain ensimmäistä pääkomponenttia saadaan havaintoja x_1, x_2,...,x_n vastaavat pääkomponenttipisteet muunnoksella (Mellin, 2004b):

$$y_{1j} = \mathbf{b}^{(1)T}\mathbf{x}_j \text{ , } j = 1,2,...,n \tag{3.3}$$

missä $\mathbf{b}^{(1)}$ on suurinta ominaisarvoa vastaava ominaisvektori.

3.2 Faktorianalyysi[9]

Faktorianalyysi muistuttaa monessa suhteessa pääkomponenttianalyysia. Faktorianalyysin suurimmat eroavaisuudet ovat muuttujien kokonaisvaihtelun jaottelu yhteis- ja ominaisvaihteluun sekä rotatointi. Rotatoinnin avulla pyritään faktoreissa, jotka vastaavat pääkomponentteja, ns. yksinkertaiseen rakenteeseen. Tälle rakenteelle pyritään antamaan tutkittavan ilmiön teoriaan liittyvä tulkinta. Faktorianalyysissa ei siis ole kyse kokonaisvaihtelun maksimaalisesta siirtämisestä uusille muuttujille, vaan vähäulotteisen piilorakenteen löytämisestä muuttujien korrelaatioiden avulla.

Faktorianalyysin juuret ulottuvat 1800-luvun loppupuolelle, jolloin biologiassa omaksuttiin ajatus latentista eli piilevästä muuttujasta, jota ei voida suoraan havaita tai mitata (Leskinen 1987, 1).

[9] Luvun tiedot perustuvat lähteeseen Mustonen (1995), ellei toisin mainittu.

Faktorianalyysin isänä pidetään yleisesti psykologi Charles Spearmania. Hänen uraauurtava työnsä älykkyyden mittaamiseksi faktorien avulla sai alkunsa vuonna 1904. Tilastotieteen näkökulmasta merkittävä rooli faktorianalyysin syntymiselle oli Karl Pearsonilla. (Harman 1976, 3)

Olkoon aineistossa p muuttujaa, $X_1,X_2,...,X_p$. Faktorianalyysin mallissa oletetaan, että on olemassa m (m<p) kappaletta piileviä taustamuuttujia eli faktoreita $F_1,F_2,...,F_m$. Yleisesti faktorianalyysissa havaitut muuttujat voidaan esittää mallin

$$\begin{aligned} X_1 &= \mu_1 + a_{11}F_1 + \cdots + a_{1m}F_m + U_1 \\ X_2 &= \mu_2 + a_{21}F_1 + \cdots + a_{2m}F_m + U_2 \\ &\vdots \quad \vdots \qquad\quad \vdots \qquad \vdots \\ X_p &= \mu_p + a_{p1}F_1 + \cdots + a_{pm}F_m + U_p \end{aligned} \tag{3.4}$$

avulla, joka kirjoitetaan usein matriisimuodossa

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{AF} + \mathbf{U}. \tag{3.5}$$

Tässä muuttujat $\mathbf{F}=(F_1,F_2,...,F_m)$ ovat yhteisfaktoreita ja muuttujat $\mathbf{U}=(U_1,U_2,...,U_p)$ ominaisfaktoreita. Faktorimatriisi

$$\mathbf{A} = \begin{bmatrix} a_{11}\, a_{12} \cdots a_{1m} \\ a_{21}\, a_{22} \cdots a_{2m} \\ \vdots \\ a_{p1}\, a_{p2} \cdots a_{pm} \end{bmatrix}$$

koostuu faktorilatauksista a_{ij}, jotka ilmaisevat miten havaitut muuttujat riippuvat faktoreista. Latausmatriisissa **A** faktorilatausten neliöiden summa vaakariveittäin on nimeltään muuttujan i kommunaliteetti. Se ilmaisee, kuinka suuren osan muuttujan i vaihtelusta faktorit yhteensä selittävät. Faktorilatausten neliöiden summaa pystyriveittäin kutsutaan faktorin j ominaisarvoksi, joka ilmaisee kuinka tärkeä faktori j on selittämään analyysissa mukana olevia havaittuja muuttujia. (Leskinen 1987, 7-9).

Satunnaisvektoreihin **X**, **F** ja **U** sekä latausmatriisiin **A** liitetään seuraavat yleiset oletukset:

1. $\mathbf{X} \sim N(\mu,\Sigma)$, $\Sigma>0$,
2. $\mathbf{F} \sim N(\mathbf{0},\Phi)$,
3. $\mathbf{U} \sim N(\mu,\Psi^2)$, missä $\Psi^2 = \operatorname{diag}(\psi_1^2, \psi_2^2,\ldots,\psi_p^2)$,
4. Ominaisfaktorit **U** ovat riippumattomia yhteisfaktoreista **F**,
5. Faktorimatriisin **A** aste on täysi eli $r(\mathbf{A}) = m$.

Esitetyt oletukset antavat kovarianssimatriisille Σ rakenteen, jota kutsutaan faktorianalyysin perusyhtälöksi ja se voidaan esittää muodossa

$$\Sigma = \mathbf{A}\Phi\mathbf{A}^T + \Psi^2. \tag{3.6}$$

Tavallisesti tehdään vielä yksinkertaistava oletus

6. $\Phi = \mathbf{I}$

eli faktorit oletetaan keskenään korreloimattomiksi, jolloin perusyhtälö on

$$\Sigma = \mathbf{A}\mathbf{A}^T + \Psi^2. \tag{3.7}$$

Tällöin matriisi **A** antaa muuttujien ja faktorien väliset kovarianssit. Jos analyysi tehdään korrelaatiomatriisista, niin $\rho(X_i,F_j) = a_{ij}$ eli faktorimatriisi **A** sisältää muuttujien ja faktorien väliset korrelaatiokertoimet.

Faktorit **F** eivät määräydy esitetyistä oletuksista ja perusyhtälöstä yksikäsitteisesti, vaan jäljelle jää mahdollisuus rotatointiin. Rotatoinnissa teknisesti yhtä hyvien ratkaisujen joukosta valitaan sellainen, joka on tulkinnallisesti selkein.

Olkoon **Q** mielivaltainen m×m-neliömatriisi[10], tällöin $\mathbf{F}^*=\mathbf{Q}^T\mathbf{F}$. Jos merkitään $\Phi^*=\mathbf{Q}^T\Phi\mathbf{Q}$ ja $\mathbf{A}^*=\mathbf{A}(\mathbf{Q}^{-1})$, saadaan muuttujille **X** esitys

$$\mathbf{X} = \mu + \mathbf{A}^*\mathbf{F}^* + \mathbf{U} \tag{3.8}$$

[10] Kirjallisuudessa matriisista **Q** käytetään usein nimitystä rotatointimatriisi.

ja perusyhtälö tulee muotoon

$$\Sigma = \mathbf{A}^*\mathbf{\Phi}^*\mathbf{A}^{*T} + \mathbf{\Psi}^2. \tag{3.9}$$

Koska faktorit **F** eivät määräydy esitetyistä oletuksista ja perusyhtälöstä yksikäsitteisesti, niin jokainen $\mathbf{F}^*=\mathbf{Q}^T\mathbf{F}$ ja sitä vastaava $\mathbf{A}^*=\mathbf{A}(\mathbf{Q}^{-1})$ on myös mahdollinen ratkaisu.

3.3 Poisson-regressio [11]

Poisson-jakauma on nimetty keksijänsä Simeon Poissonin mukaan. Poisson todisti, että binomijakauma Bin(n,p) lähestyy Poisson-jakaumaa Poisson(λ) kun $n \rightarrow \infty$, jos $p = \frac{\lambda}{n}$ ja $\lambda > 0$. Näiden oletusten täyttyessä satunnaismuuttujan Y voidaan todeta noudattavan Poisson-jakaumaa parametrilla λ, ja sen todennäköisyysfunktio voidaan esittää muodossa

$$P(Y = y) = \begin{cases} \dfrac{\lambda^y}{y!}e^{-\lambda}, & y = 0,1,2,..., \\ 0 & \text{muualla.} \end{cases} \tag{3.10}$$

Poisson-jakauma on käyttökelpoinen kun mallinnetaan harvinaisia tapahtumia tai tilanteita, joissa selitettävä muuttuja saa pieniä ei-negatiivisia kokonaislukuarvoja. Jakauman ominaisuus on, että muuttujan odotusarvo ja varianssi ovat yhtä suuria. (Long, 1997)

Havaintoaineistossa $x_{1i},...,x_{pi}$, y_i, $i=1,2,...,n$, on y_i Poissonin jakaumaa noudattavan muuttujan Y_i arvo. Odotusarvo $E[Y_i] = \lambda_i$ riippuu selittävien muuttujien X_j arvoista ja

$$P(Y_i = y) = \frac{\lambda_i^y}{y!}e^{-\lambda_i} \quad , y = 0,1,2,..., \text{ja } \lambda > 0. \tag{3.11}$$

Y_i:n varianssi on $Var[Y_i]= \lambda_i$. Varianssi siis riippuu selitettävän muuttujan odotusarvosta, joten ei ole hyviä edellytyksiä pienimmän neliösumman regressiolle.

---

[11] Luku perustuu kokonaisuudessaan lähteeseen Laininen (2000), josta löytyy tarkempaa tietoa jakaumasta, parametrien estimoinnista ja sovelluskohteista.

Eräs vaihtoehto on mallintaa satunnaismuuttuja Y lineaarisen regression avulla eli

$$E[Y] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p = \mathbf{x}^T\boldsymbol{\beta}. \tag{3.12}$$

Tämä malli voi antaa myös negatiivisia arvoja, jotka eivät ole Y:lle mahdollisia. Usein käytetäänkin mallia

$$E[Y] = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p) = \exp(\mathbf{x}^T\boldsymbol{\beta}). \tag{3.13}$$

Koska $E[Y_i] = \lambda_i$ voidaan malli kirjoittaa muotoon

$$\lambda_i = \exp(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}),\ i = 1,2,...,n \tag{3.14}$$

joka on yhtäpitävä esityksen

$$\ln\lambda_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} = \mathbf{x}^T\boldsymbol{\beta} \tag{3.15}$$

kanssa. Tämä log-lineaarinen malli on yleisin tapa esittää muuttuja λ_i (Greene, 2000).

Mallin parametrien estimointia ei siis voida suorittaa PNS-menetelmällä, vaan ne estimoidaan suurimman uskottavuuden menetelmällä. Koska uskottavuusfunktion maksimikohdan määräävä osittaisderivaattayhtälöryhmä on epälineaarinen, täytyy regressiokertoimien estimaatit hakea iteratiivisella menetelmällä. Tavallisesti käytetään iteratiivista painotettua pienimmän neliösumman menetelmää, jonka avulla saadaan selville myös estimaattien varianssit ja kovarianssit. (Neter et al. 1996; ks. Laininen 2000, 151)

3.4 Negatiivinen binomijakauma –regressio [12]

Negatiivinen binomijakauma –regressio on Poisson-regression yleistys. Poisson-jakauman odotusarvo ja varianssi ovat yhtäsuuria. Käytännössä varianssi on kuitenkin usein odotusarvoa suurempi ja tällaista ilmiötä kutsutaan ylihajonnaksi. Ylihajonta on eräs merkki siitä, että

[12] Luku perustuu kokonaisuudessaan lähteeseen Long (1997), josta löytyy tarkempaa tietoa jakaumasta, parametrien estimoinnista ja sovelluskohteista.

havaintojen ei voida olettaa olevan peräisin Poisson-jakaumasta. Kirjallisuudessa on esitetty Poisson-jakaumaan lukuisia muunnoksia, joiden tarkoituksena on ollut korjata havaittua ylihajontaa.

Negatiivinen binomijakauma –regressio voidaan määrittää muuttamalla edellisessä luvussa esitetty Poisson-jakauman odotusarvo (3.15) muotoon (Greene, 2000):

$$\ln \lambda_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \varepsilon_i,\ i = 1,2,\ldots,n\,. \qquad (3.16)$$

Yleensä oletetaan, että $\exp(\varepsilon_i)$ noudattaa Gamma-jakaumaa tuottaen ehdollisen todennäköisyyden (Greene, 2000):

$$P(Y = y_i \mid \varepsilon) = \frac{\exp(-\lambda_i)\exp(\varepsilon)\lambda_i^{y_i}}{y_i!}. \qquad (3.17)$$

Integroimalla ε yhtälöstä pois, saadaan negatiivisen jakauman todennäköisyysfunktioksi (funktion yksityiskohtainen johtaminen esim. Greene, 2000):

$$P(Y = y_i) = \frac{\Gamma(\varphi + y_i)}{\Gamma(\varphi) y_i!\, u_i^{\varphi} (1 - u_i)^{y_i}},\ y = 0,1,2,\ldots, \qquad (3.18)$$

missä

$\Gamma(.)$ = Gamma funktio

φ = $1/\alpha$, jossa α on ylihajontaparametri

u_i = $\varphi / (\varphi + \lambda_i)$

$\ln \lambda_i$ = $\mathbf{x}^T \boldsymbol{\beta}$.

Tällöin satunnaismuuttujan Y odotusarvo on sama kuin Poisson-jakauman tapauksessa, mutta varianssi eroaa odotusarvosta, sillä

$$\mathrm{Var}[Y_i] = E[Y_i](1 + \alpha\, E[Y_i]). \qquad (3.19)$$

Negatiivinen binomijakauma poikkeaa Poisson-jakaumasta ainoastaan parametrin α osalta, joka korjaa havaintojen ylihajontaa. Varianssin kaavasta nähdään, että negatiivisen binomijakauman malli pelkistyy Poisson-malliksi jos ylihajontaa mittaava parametri α=0. Negatiivisen binomijakauman parametrit estimoidaan suurimman uskottavuuden menetelmällä. (Long, 1997)

3.5 BVDIP-regressio

Karlis & Ntzoufras (2003) esittivät aiempiin urheiluvedonlyönnin tutkimuksiin mielenkiintoisen lisäyksen, jossa tasapelin todennäköisyyttä kasvatetaan. He kutsuivat malliaan BVDIP-regressioksi. Ajatellaan satunnaismuuttujien X_k k=1,2,3 noudattavan Poisson-jakaumaa parametrein $\lambda_1,\lambda_2>0$ ja $\lambda_3\geq 0$. Tällöin satunnaismuuttujat $X=X_1+X_3$ ja $Y=X_2+X_3$ noudattavat kaksiulotteista Poisson-jakaumaa BivPoisson($\lambda_1,\lambda_2,\lambda_3$) todennäköisyysfunktiolla

$$P_{X,Y}(x,y) = P(X=x, Y=y) = \exp\{-(\lambda_1+\lambda_2+\lambda_3)\}\frac{\lambda_1^x}{x!}\frac{\lambda_2^y}{y!}\sum_{k=0}^{\min(x,y)}\binom{x}{k}\binom{y}{k}k!\left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^k, \quad (3.20)$$

missä

$$\ln\lambda_{1i} = \beta_{10}+\beta_{11}x_{1i}+\cdots+\beta_{1p}x_{pi} = \mathbf{x}^T\boldsymbol{\beta}_1$$

$$\ln\lambda_{2i} = \beta_{20}+\beta_{21}x_{1i}+\cdots+\beta_{2p}x_{pi} = \mathbf{x}^T\boldsymbol{\beta}_2$$

$$\ln\lambda_{3i} = L.$$

Tässä kahden muuttujan yhteisjakaumassa satunnaismuuttujien X ja Y ei tarvitse olla riippumattomia. Muuttujien reunajakaumat ovat Poisson-jakautuneita odotusarvoilla $E(X)=\lambda_1+\lambda_3$ ja $E(Y)=\lambda_2+\lambda_3$. Muuttujien välinen kovarianssi $cov(X,Y)=\lambda_3$. Jos $\lambda_3=0$ niin muuttujat ovat riippumattomia ja niiden yhteisjakauma pelkistyy kahden Poisson-jakautuneen muuttujan tuloksi. Kaksiulotteisen Poisson-jakauman parametrien estimoimiseksi on olemassa useita menetelmiä. Työssään Karlis & Ntzoufras (2003) esittivät EM-algoritmin[13] käyttöä, joka on tehokas menetelmä suurimman uskottavuuden estimaattien määrittämiseksi.

Useissa aiemmissa tutkimuksissa on todettu tilastollisten mallien aliarvioivan tasapelien todennäköisyyttä. BVDIP-regressiossa tilannetta korjataan kasvattamalla todennäköisyysmatriisin **P** diagonaalin alkioiden todennäköisyyttä. Tällöin BVDIP-mallin todennäköisyysfunktio voidaan esittää muodossa

---

[13] Tarkempaa lisätietoa EM-algoritmista (Expectation-Maximization) esim. McLachlan & Krishnan (1997).

$$\mathrm{BVDIP}(x,y)=\begin{cases}(1-p)\mathrm{BivPoisson}(x,y\mid\lambda_1,\lambda_2,\lambda_3), & x\neq y\\ (1-p)\mathrm{BivPoisson}(x,y\mid\lambda_1,\lambda_2,\lambda_3)+pf_D(x\mid\theta), & x=y\end{cases} \tag{3.21}$$

missä $f_D(x|\theta)$ on diagonaalille sovitettu diskreetti todennäköisyysjakauma. Mahdollisia vaihtoehtoja tällaisiksi ovat esimerkiksi Poisson-jakauma tai Geometrinen-jakauma. Tällöin muuttujien reunajakaumat eivät ole enää Poisson-jakaumia, vaan niissä yhdistyvät sekä Poisson-jakauma että diagonaalille määrätty jakauma. Termi p ilmoittaa sekoitussuhteen, jolla nämä kaksi komponenttia on yhdistetty. (Karlis & Ntzoufras, 2003)

Tässä työssä todennäköisyysmatriisin **P** diagonaalille sovitetuksi jakaumaksi otetaan yksinkertainen diskreettijakauma Disc(J)[14], jonka todennäköisyysfunktio on (Karlis & Ntzoufras, 2003):

$$f(x\mid\theta,J)=\begin{cases}\theta_x & \text{kun } x=0,1,2,\ldots,J\\ 0 & \text{muualla}\end{cases} \tag{3.22}$$

siten, että $\sum_{x=0}^{J}\theta_x=1$. Jos J=0 pelkistyy malli ZIP-regressioksi[15], sillä tällöin $\theta_0=1$. Tässä tapauksessa siis ainoastaan tuloksen 0-0 todennäköisyyttä kasvatettaisiin arvolla p eli sekoitussuhteen verran.

---

[14] Tutkimuksissaan Karlis & Ntzoufras totesivat jakauman Disc(3) parhaaksi eri jalkapallosarjojen mallintamiseen, joten tässä työssä on käytetty diagonaalin kasvattamiseen samaa jakaumaa. Jääkiekkosarjoissa maalimäärän odotusarvo on suurempi, ja niiden osalta diagonaalin kasvattamiseen on käytetty Poisson-jakaumaa.

[15] ZIP eli Zero Inflated Poisson. Sopii jakaumaksi aineistoihin, joissa 0 havaintojen lukumäärä on suurempi kuin muuttujan odotusarvon perusteella voitaisiin olettaa. Lisätietoja jakaumasta ja sen estimoinnista esim. Long (1997).

4. Ennustemallin konstruointi

Tässä luvussa rakennetaan ennustemallit 1X2-todennäköisyysarvioiden laskemiseksi. Luku alkaa maalimäärien jakaumaoletusten testaamisella. Sen jälkeen esitellään pelisysteemin tietokanta ja sen muuttujat, joista faktorianalyysin ja pääkomponenttianalyysin avulla etsitään parhaat selittävät muuttujat koti- ja vierasmaalimäärien odotusarvojen laskemiseksi.

Luvun esimerkkisarjana toimii Veikkausliiga[16]. Aineistona luvuissa 4.1 ja 4.2 on 842 Veikkausliigan ottelua kausilta 1999-2003. Luvusta 4.4 eteenpäin aineistona on 221 ottelua Veikkausliigasta kausilta 2002-2003. Luvussa 4.3 aineistona käytetään poikkeuksellisesti Englannin Valioliigan otteluita. Tämä johtuu siitä, että tässä työssä esitettävää maalimatriisin korjaustermiä voidaan verrata aiempien tutkimusten menetelmiin.

4.1 Jakaumaoletusten testaus

4.1.1 Poisson-jakauma

Kausilta 1999-2003 on Veikkausliigasta käytettävissä 842 ottelun tulokset. Kaudella 2002 käytössä oli hieman erilainen sarjasysteemi, joten kyseiseltä kaudelta on jätetty huomioimatta ala- ja yläloppusarjojen ottelut. Taulukkoon 2 on koottu aineiston tunnuslukuja.

Taulukko 2: Koti- ja vierasjoukkueen maalimäärien tunnuslukuja Veikkausliigasta kausilta 1999-2003.

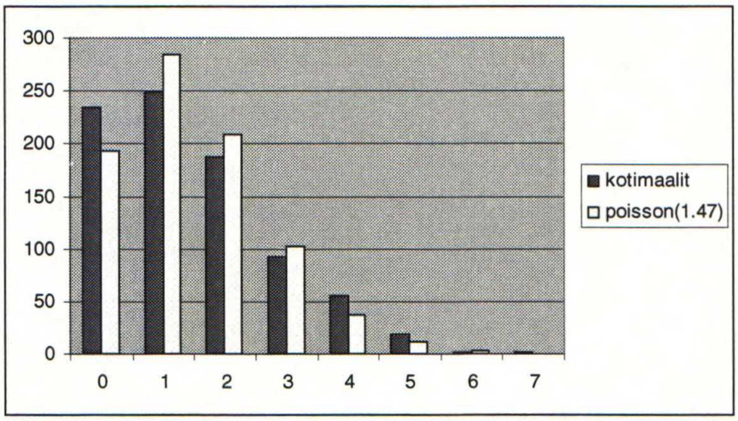
	keskiarvo	mediaani	moodi	varianssi	minimi	maksimi	summa	havaintoja
kotijoukkue	1.47	1	1	1.77	0	7	1237	842
vierasjoukkue	1.15	1	0	1.38	0	7	968	842

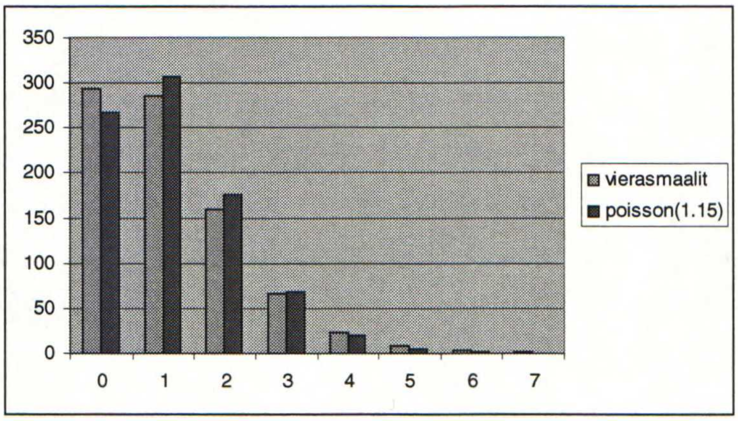
Kuten taulukosta nähdään, tekee kotijoukkue satunnaisessa ottelussa keskimäärin 1,47 maalia ja vierasjoukkue vastaavasti 1,15 maalia. Maalimäärien erotus 0,32 maalia voidaan tulkita kotieduksi. Taulukosta huomataan myös, että maalien varianssi on sekä koti- että vierasjoukkueen kohdalla suurempi kuin maalien keskiarvo.

Kuvassa 1 on esitetty koti- ja vierasjoukkueen maalimäärien havaitut frekvenssit sekä niille lasketut Poisson- jakauman mukaiset teoreettiset frekvenssit. Seuraavaksi testataan, voidaanko oletus maalimäärien Poisson-jakautuneisuudesta hyväksyä.

[16] Veikkausliiga on Suomen korkein sarjataso jalkapallossa.

Kuva 1: Koti- ja vierasmaalien jakaumat sekä keskiarvolla estimoitu Poisson-jakauma.





Maalimäärien Poisson-jakautuneisuuden testaamiseen käytetään χ^2- yhteensopivuustestiä. Sen teorian esitti Karl Pearson vuonna 1900. Se on ensimmäinen varsinainen todennäköisyyksiin perustuva tilastollinen testi ja sen keksimistä on pidettu modernin tilastotieteen alkuna. (Laininen 2004, 48)

Käytetään χ^2- yhteensopivuustestiä Poisson-jakautuneisuuden tutkimiseen, jolloin testattavat hypoteesit ovat:

H_0: maalimäärät ovat Poisson-jakautuneita

H_1: maalimäärien jakauma ei ole Poisson-jakauma.

Testisuure on tällöin:

$$\chi^2 = \sum_{j=1}^{k} \frac{(f_i - e_i)^2}{e_i}, \tag{4.1}$$

missä f_i on havaittu frekvenssi ja e_i on Poisson-jakauman mukainen odotettu frekvenssi. Testi noudattaa likimain jakaumaa $\chi^2(k\text{-}r\text{-}1)$, missä k on havaittujen luokkien määrä ja r on odotettujen frekvenssien laskemiseksi estimoitujen parametrien määrä.

χ^2- yhteensopivuustestiin liitetään yleensä seuraavat oletukset:

1. kukin havainto ja frekvenssi f_i on riippumaton muista havainnoista
2. havaintojen lukumäärä $n \geq 50$
3. kaikki odotetut frekvenssit $e \geq 1$
4. korkeintaan 20 % odotetuista frekvensseistä ≤ 5.

Kuvasta 1 havaittiin, että koti- ja vierasmaaleissa kaksi viimeistä luokkaa ovat havaituilta frekvensseiltä erittäin pieniä. Keskiarvon perusteella voidaan päätellä, että myös niiden odotetut frekvenssit tulevat olemaan pieniä. Jotta testin oletukset täyttyvät, yhdistellään maalimäärien siten, että kotimaaleille suurin luokka on 6+ maalia ja vierasmaaleille vastaavasti 5+ maalia. Yhdistämisen jälkeen testi voidaan suorittaa ja sen tulokset on esitetty taulukossa 3.

Taulukko 3: Tulokset maalimäärien Poisson-jakautuneisuuden testistä.

KOTIMAALIT			VIERASMAALIT		
maalien lkm	**havaitut**	**poisson**	**maalien lkm**	**havaitut**	**poisson**
0	235	193.8	0	293	266.7
1	249	284.7	1	286	306.6
2	188	209.1	2	160	176.2
3	93	102.4	3	67	67.5
4	56	37.6	4	24	19.4
5	18	11.1	5+	12	4.5
6+	3	2.7			
keskiarvo	1.47		keskiarvo	1.15	
χ^2	29.78		χ^2	19.29	
df	5		df	4	
p-arvo	0.00002		p-arvo	0.0007	

Nollahypoteesit joudutaan hylkäämään sekä koti- että vierasmaalien osalta. Testin tulos on linjassa aiempien tutkimusten kanssa. Vaikka jakaumien visuaalinen muoto ja lajin luonne itsessään viittaavaat Poisson-jakaumaan, eivät tilastolliset testit yleensä anna kovinkaan vahvaa tukea oletukselle Poisson-jakautuneisuudesta.

4.1.2 Negatiivinen binomijakauma

Tutkitaan seuraavaksi noudattavatko maalimäärät negatiivista binomijakaumaa. Testattavat hypoteesit ovat:

H_0: maalimäärät noudattavat negatiivista binomijakaumaa

H_1: maalimäärät eivät noudata negatiivista binomijakaumaa.

Testin tulokset selviävät taulukosta 4.

Taulukko 4: Tulokset maalimäärien negatiivisen binomijakautuneisuuden testistä.

KOTIMAALIT			VIERASMAALIT		
maalien lkm	**havaitut**	**NegBin**	**maalien lkm**	**havaitut**	**NegBin**
0	235	221.0	0	293	295.1
1	249	269.8	1	286	282.8
2	188	187.5	2	160	159.1
3	93	97.5	3	67	68.5
4	56	42.2	4	24	25.0
5	18	16.0	5+	12	8.1
6+	3	5.5			
keskiarvo	1.47		keskiarvo	1.15	
varianssi	1.77		varianssi	1.38	
χ^2	8.64		χ^2	1.98	
df	5		df	4	
p-arvo	0.124		p-arvo	0.739	

Tuloksista nähdään, että negatiivinen binomijakauma sopii kohtuullisen hyvin maalimäärien mallintamiseen. Nollahypoteesit voidaan hyväksyä sekä kotimaalien että vierasmaalien osalta. Jotta Poisson-regressiolla ja NegBin-regressiolla voidaan laskea arviot ottelun eri lopputuloksille, oletetaan koti- ja vierasmaalit riippumattomiksi. Seuraavaksi tutkitaan tämän riippumattomuusoletuksen paikkansapitävyyttä.

4.2 Riippumattomuuden testaus

Edellisessä luvussa esitetyn χ^2-yhteensopivuustestillä voidaan tutkia myös kahden satunnaismuuttujan X ja Y välistä riippumattomuutta. Yhteisjakaumasta poimitaan n havaintoa, jotka luokitellaan sekä X:n että Y:n mukaan. X:n luokkien frekvenssejä merkitään $n_{i.}$ (i =1,2,...,r) ja Y:n frekvenssejä $n_{.j}$ (j =1,2,...,c). (Laininen 2004, 53)

Jos muuttujat X ja Y ovat riippumattomia, solujen ij odotetut frekvenssit e_{ij} saadaan lasketuksi

$$e_{ij} = \frac{n_{i.} n_{.j}}{n}. \tag{4.2}$$

Testisuure on tässä tapauksessa

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \tag{4.3}$$

$\chi^2 \sim \chi^2((r-1)(c-1))$. Testiä koskevat lisäksi edellisessä luvussa esitetyt oletukset. (Laininen 2004, 54)

Jotta testin soveltamisehdot täyttyvät, joudutaan joukkueiden maalilukuja jälleen yhdistämään siten, että kotijoukkueen ja vierasjoukkueen suurin luokka on 4+ maalia. Testattavat hypoteesit ovat:

H_0: koti- ja vierasmaalit ovat riippumattomia

H_1: koti- ja vierasmaalit ovat riippuvia.

Testin tulokset on esitetty koostusti taulukoissa 5 ja 6.

Taulukko 5: Ottelun tulosten havaitut frekvenssit (suluissa odotetut frekvenssit).

	VIERASMAALIT				
KOTIMAALIT	0	1	2	3	4+
0	76 (81.8)	81 (79.8)	40 (44.7)	26 (18.7)	12 (10.0)
1	81 (86.6)	91 (84.6)	45 (47.3)	20 (19.8)	12 (10.6)
2	68 (65.4)	55 (63.9)	41 (35.7)	14 (15.0)	10 (8.0)
3	37 (32.4)	32 (31.6)	18 (17.7)	5 (7.4)	1 (4.0)
4+	31 (26.8)	27 (26.2)	16 (14.6)	2 (6.1)	1 (3.3)

Taulukko 6: Riippumattomuustestin solukohtaiset arvot ja testin tulokset.

	VIERASMAALIT				
KOTIMAALIT	0	1	2	3	4+
0	0.408	0.017	0.485	2.850	0.379
1	0.368	0.488	0.113	0.002	0.172
2	0.102	1.229	0.779	0.062	0.479
3	0.665	0.005	0.006	0.779	2.228
4+	0.660	0.027	0.128	2.780	1.596
χ^2	16.81				
df	16				
p-arvo	0.398				

Testi antaa tukea nollahypoteesille, joten oletus maalimäärien riippumattomuudesta voidaan hyväksyä. Vaikka useissa tutkimuksissa koti- ja vierasmaalit on todettu tilastollisesti riippumattomiksi, poikkeavat jotkut yksittäiset tulokset riippumattomuusoletuksesta. Tätä ilmiötä on kirjallisuudessa selitetty peliin liittyvillä ehdollisilla tilanteilla. Jos ottelu etenee lähelle peliajan loppua maalittomassa tasatilanteessa, saattavat joukkueet varmistella tasapelistä saatavaa pistettä, mikä voi johtaa odotettua suurempaan 0-0 tuloksen frekvenssiin. Toisinaan toinen joukkueista saattaa saada pienoisen onnen avustuksella muutaman maalin johdon. Tällöin häviöllä oleva joukkue saattaa hyväksyä tappion ja alkaa säästää itseään seuraavaan peliin. Tällainen tilanne saattaa johtaa runsasmaalisten koti- ja vierasvoittojen odotettua suurempaan frekvenssiin.

BVDIP-regressiossa joidenkin tulosten mahdollinen riippuvuus ei ole ongelma, koska mallin parametri λ_3 korjaa tilannetta. Sen sijaan Poisson-regressiossa ja NegBin-regressioissa mahdollinen riippuvuus aiheuttaa ongelmia. Tämän takia niiden todennäköisyysmatriisia **P** on korjattava ennen lopullisten todennäköisyyksien laskemista. Seuraavassa luvussa käsitellään tällaisen korjaustermin määrittämistä.

4.3 Tulosmatriisin korjaustermi

Jotta työssä esitettävän korjaustermin kyvykkyyttä voidaan verrata muihin menetelmiin, käytetään tässä luvussa Raitasen (1999) tutkielman aineistoa Englannin Valioliigasta kausilta 1994-1998. Yhteensä aineisto käsittää 1602 ottelua ja sen maalimatriisi **M** on esitetty taulukossa 7.

Taulukko 7: Maalimatriisi M Englannin Valioliigasta kausilta 1994-1998, yhteensä 1602 ottelua.

	VIERASMAALIT				
KOTIMAALIT	0	1	2	3	4+
0	152	118	69	26	12
1	190	200	81	50	15
2	120	133	73	31	12
3	68	70	29	20	5
4	26	26	22	12	1
5+	19	13	7	2	0
keskiarvo(h)	1.516				
keskiarvo(a)	1.089				

Aineistossa on yhdistetty kotimaaleille suurimmaksi luokaksi 5+ maalia ja vierasmaaleille 4+ maalia. Keskimäärin kotijoukkue on tehnyt 1.516 maalia ottelussa ja vierasjoukkue 1.089 maalia.

Matriisista **M** laskettu odotettujen frekvenssien matriisi **E** on esitetty taulukossa 8 ja χ^2-testisuureen solukohtaiset arvot ja riippumattomuustestin tulokset taulukossa 9.

Taulukko 8: Aineiston odotettujen frekvenssien matriisi E.

	VIERASMAALIT				
KOTIMAALIT	0	1	2	3	4+
0	135.3	131.8	66.1	33.2	10.6
1	192.4	187.4	94.0	47.2	15.1
2	132.4	129.0	64.7	32.5	10.4
3	68.9	67.1	33.7	16.9	5.4
4	31.2	30.4	15.3	7.7	2.4
5+	14.7	14.3	7.2	3.6	1.2

Taulukko 9: Riippumattomuustestin solukohtaiset arvot ja testin tulokset.

	VIERASMAALIT				
KOTIMAALIT	0	1	2	3	4+
0	2.057	1.442	0.125	1.554	0.188
1	0.030	0.852	1.802	0.169	0.000
2	1.169	0.125	1.058	0.067	0.258
3	0.012	0.124	0.650	0.569	0.029
4	0.875	0.640	2.977	2.463	0.853
5+	1.247	0.124	0.005	0.717	1.152
χ^2	23.33				
df	20				
p-arvo	0.273				

Taulukosta 9 nähdään, että joidenkin tulosten χ^2-suureet saavat huomattavan suuria arvoja. Tämä on merkki siitä, että näiden tulosten kohdalla riippumattomuusoletus on kyseenalainen ja siksi näiden ongelmallisten tulosten odotettuja todennäköisyyksiä täytyy korjata.

Kirjallisuudessa on esitetty joitakin korjaustermejä ongelmallisten tulosten korjaamiseksi. Dixon & Coles (1997) esittivät menetelmän, jossa tasapelien 0-0 ja 1-1 todennäköisyyttä kasvatetaan tulosten 0-1 ja 1-0 kustannuksella. Tämä menetelmä ei kuitenkaan toimi yleisesti, koska muilla tuloksilla koti- ja vierasmaalien reunatodennäköisyydet muuttuvat ja todennäköisyyksien summasta tulee erisuuri kuin yksi. Hyvälle korjausmatriisille voidaan asettaa kaksi vaatimusta:

1. matriisin avulla on mahdollista korjata vain valittuja tuloksia
2. matriisi ei saa muuttaa reunajakaumien todennäköisyyksiä.

Ensimmäinen vaatimus on seurausta siitä, että läheskään kaikki tulokset eivät ole ongelmallisia. Koska suurten maalimäärien tuloksissa (h,a >3) vaihtelu on suurta, saatetaan tällaisten tulosten korjaustermillä jopa huonontaa tilannetta[17]. Toinen vaatimus edellyttää reunajakaumien muuttumattomuutta, jolloin kaikkien tulosten todennäköisyydet summautuvat yhteen.

Raitasen (1999) esittämä menetelmä täyttää molemmat asetetut vaatimukset. Se perustuu sarjakohtaisen ns. riippuvuusparametrin määrittämisen, jonka avulla voidaan määrittää korjaustermit valituille tuloksille. Määrittämisprosessi on iteratiivinen ja siinä riippuvuusparametrin arvoa kasvatetaan vakiomäärällä jokaisella iteraatiokierroksella, kunnes χ^2-riippumattomuustestin testisuure ei enää pienene. Raitanen korjasi pahiten taulukossa 9 riippumattomuudesta poikenneita tuloksia 0-0, 2-0, 2-1 ja 0-1. Hänen määrittämänsä korjausmatriisi on esitetty taulukossa 10.

Taulukko 10: Raitasen määrittämä korjausmatriisi Valioliigan aineistolle.

	VIERASMAALIT				
KOTIMAALIT	0	1	2	3	4+
0	1.117	0.880	1.000	1.000	1.000
1	1.000	1.000	1.000	1.000	1.000
2	0.880	1.123	1.000	1.000	1.000
3	1.000	1.000	1.000	1.000	1.000
4	1.000	1.000	1.000	1.000	1.000
5+	1.000	1.000	1.000	1.000	1.000

Tässä työssä esitetään uusi menetelmä korjausmatriisin määrittämiseksi. Menetelmä perustuu siihen, että minimoidaan χ^2-riippumattomuustestin testisuuretta muuttamalla valittujen tulosten korjausmatriisin alkioita. Määritetään matriisit **M** = havaittu maalimatriisi, **E** = maalimatriisista laskettu odotettujen frekvenssien matriisi ja $\mathbf{C}_{\text{sarja S}}$ = sarjaan S määritettävä odotettujen frekvenssien korjausmatriisi.

$$\mathbf{M} = \begin{pmatrix} \mathbf{m}_{00} & \mathbf{m}_{01} \cdots & \mathbf{m}_{0a_{max}} \\ \vdots & \vdots & \vdots \\ \mathbf{m}_{h_{max}0} & \cdots & \mathbf{m}_{h_{max}a_{max}} \end{pmatrix}$$

[17] Esimerkiksi Valioliigassa yhden kauden aikana odotettu lukumäärä tulokseen 4-3 päättyville otteluille on noin yksi. Jonain kautena tällaisia harvinaisia otteluita saattaa tulla esimerkiksi kolme tai neljä kappaletta. Jos seuraavien kausien osalta odotettuja frekvenssejä korjataan tällaisen poikkeuksellisen aineiston avulla, ovat ennusteet systemaattisesti pahasti väärässä.

$$\mathbf{E}=\begin{pmatrix}\mathbf{e}_{00} & \mathbf{e}_{01}\cdots & \mathbf{e}_{0a_{max}}\\ \vdots & \vdots & \vdots\\ \mathbf{e}_{h_{max}0} & \cdots & \mathbf{e}_{h_{max}a_{max}}\end{pmatrix}$$

$$\mathbf{C}=\begin{pmatrix}\mathbf{c}_{00} & \mathbf{c}_{01}\cdots & \mathbf{c}_{0a_{max}}\\ \vdots & \vdots & \vdots\\ \mathbf{c}_{h_{max}0} & \cdots & \mathbf{c}_{h_{max}a_{max}}\end{pmatrix}$$

Lähtötilanteessa $c_{ha} = 1$, kaikille alkioille c_{ha}. Valitsemalla haluttujen tulosten korjaustermit c_{ha} kohdemuuttujiksi voidaan minimointi kirjoittaa muotoon

$$\min_{c_{ha}} \sum_{h=0}^{h_{max}}\sum_{a=0}^{a_{max}} \frac{(m_{ha}-c_{ha}e_{ha})^2}{c_{ha}e_{ha}}$$

$$\text{s.t.} \quad \sum_{h=0}^{h_{max}} c_{ha}e_{ha} = \sum_{h=0}^{h_{max}} e_{ha}, \quad a = 0,1,\dots,a_{max}$$

$$\sum_{a=0}^{a_{max}} c_{ha}e_{ha} = \sum_{a=0}^{a_{max}} e_{ha}, \quad h = 0,1,\dots,h_{max}$$

$$c_{ha} \geq 0, \qquad h = 0,1,\dots,h_{max} \quad a = 0,1,\dots,a_{max}.$$

Verrattaessa esitettyä menetelmää Raitasen korjausmatriisiin, valitaan kohdemuuttujiksi Raitasen käyttämiä tuloksia vastaavat elementit eli c_{00}, c_{20}, c_{21} ja c_{01}. Optimoimalla määritetty korjausmatriisi $C_{Valioliiga}$ on esitetty taulukossa 11.

Taulukko 11: Optimointimenetelmällä määritetty korjausmatriisi Valioliigan aineistolle.

	VIERASMAALIT				
KOTIMAALIT	0	1	2	3	4+
0	1.088	0.910	1.000	1.000	1.000
1	1.000	1.000	1.000	1.000	1.000
2	0.910	1.092	1.000	1.000	1.000
3	1.000	1.000	1.000	1.000	1.000
4	1.000	1.000	1.000	1.000	1.000
5+	1.000	1.000	1.000	1.000	1.000

Eri menetelmien välistä kyvykkyyttä voidaan tutkia laskemalla χ^2-riippumattomuustestin tulokset korjausmatriiseilla korjatuille frekvenseille. Tämä tapahtuu laskemalla karteesinen tulo korjausmatriisien ja odotettujen frekvenssien matriisin välillä ($\mathbf{C}_{Valioliiga} \times \mathbf{E}$) ja vertaamalla korjattuja frekvenssejä havaittuihin. Vertailun tulokset on esitetty taulukossa 12.

Taulukko 12: Tulokset eri korjausmenetelmien vertailusta Valioliigan aineistolla.

	Alkuperäinen	Raitanen	Optimointi
χ^2	23.33	19.65	19.17
p-arvo	0.273	0.480	0.511

Alkuperäisessä aineistossa riippumattomuustestin testisuure oli 23.33, joka on samansuuntainen tämän työn esimerkkiaineiston testisuureen kanssa. Raitasen korjausmatriisilla testisuureen arvo saatiin tiputettua arvoon 19.65, jolloin p-arvo lähes kaksinkertaistui. Optimoimalla saatiin p-arvoksi 0.511.

Tässä työssä ei olla kiinnostuneita yksittäisten tulosten todennäköisyyksistä, vaan niiden pohjalta lasketuista 1X2-todennäköisyyksistä. Tästä syystä korjausmatriisia käytetään tasaamaan todennäköisyysmassaa yleisimpien maalilukujen alueella. Jalkapallossa koti- ja vierasmaalilukujen h,a $\leq$ 3 eri kombinaatiot kattavat noin 90 % havainnoista. Nämä 16 tuloskombinaatiota valittiin optimointimallin kohdemuuttujiksi. Optimoimalla laskettu Veikkausliigan korjausmatriisi $\mathbf{C}_{Veikkausliiga}$ on esitetty taulukossa 13.

Taulukko 13: Optimointimenetelmän tuottama korjausmatriisi Veikkausliigaan.

	VIERASMAALIT				
KOTIMAALIT	0	1	2	3	4+
0	0.951	1.024	0.911	1.328	1.000
1	0.953	1.082	0.963	0.944	1.000
2	1.063	0.872	1.165	0.876	1.000
3	1.123	0.980	0.991	0.571	1.000
4+	1.000	1.000	1.000	1.000	1.000

Korjausmatriisilla $\mathbf{C}_{Veikkausliiga}$ korjatut odotetut frekvenssit ja havaitut frekvenssit sekä riippumattomuustestin testisuure ja vastaava p-arvo on esitetty taulukossa 14.

Taulukko 14: Havaitut ja (korjatut) frekvenssit sekä testistatistiikka.

	VIERASMAALIT				
KOTIMAALIT	0	1	2	3	4+
0	76 (77.7)	81 (81.7)	40 (40.7)	26 (24.8)	12 (10.0)
1	81 (82.6)	91 (91.5)	45 (45.6)	20 (18.7)	12 (10.6)
2	68 (69.5)	55 (55.7)	41 (41.6)	14 (13.1)	10 (8.0)
3	37 (36.3)	32 (30.9)	18 (17.5)	5 (4.2)	1 (4.0)
4+	31 (26.8)	27 (26.2)	16 (14.6)	2 (6.1)	1 (3.3)

	Alkuperäinen	Optimointi
χ^2	16.81	9.01
df	16	16
p-arvo	0.398	0.913

Tuloksista nähdään, että korjauksen jälkeen koti- ja vieramaalien riippumattomuus on lähes täydellinen. Oleellisinta on kuitenkin, että korjausmatriisin avulla 1X2-todennäköisyyksien suhdetta saatiin muutettua. Tällöin malli huomioi tasapelien odotettua korkeamman todennäköisyyden.

4.4 Muuttujien ja tietokannan esittely

Internetin kautta eri maiden jalkapallosarjoista on saatavissa runsaasti erilaista tietoa. Saatavilla on otteluiden tulokset, mutta lisäksi on saatavilla yksityiskohtaisia raportteja, joissa ottelut on tilastoitu ja analysoitu yksityiskohtaisesti. Näiden suoraan mitattavissa olevien tekijöiden lisäksi yksittäisen joukkueen peliesitykseen vaikuttavat myös monet ulkopuoliset tekijät: uusien pelaajien hankinnat, loukkaantumiset, motivaatiotekijät ja valmentajavaihdokset. Tietoa on tarjolla paljon, jolloin ongelmaksi muodostuu sen järkevä ja tehokas formalisointi. Kaiken oleellisen tiedon muuttaminen numeroiksi on mahdotonta. Toisaalta jokaisen ottelun kvalitatiivinen analyysi veisi helposti useita tunteja päivässä.

Tässä työssä käytössä on tietokanta, johon on sarjakohtaisesti tallennettu taulukon 15 muuttujat.

Taulukko 15: Tietokannan sisältämät muuttujat otteluista[18].

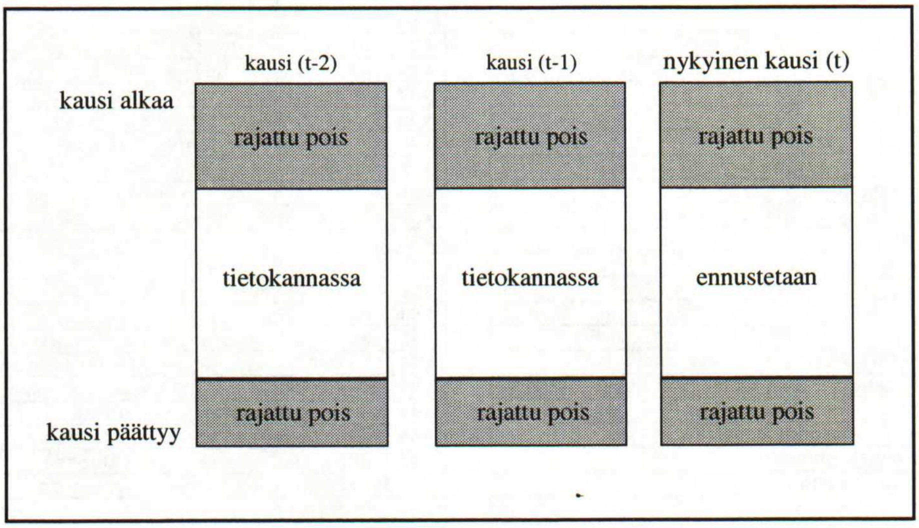
MJA	SELITE	MJA	SELITE
x1	kotimaalit	x30	kotijoukkue pisteitä / ottelu kuntopuntarissa
x2	vierasmaalit	x31	vierasjoukkue otteluita kotona
x3	kotijoukkue otteluita kotona	x32	vierasjoukkueen voitto-% kotona
x4	kotijoukkueen voitto-% kotona	x33	vierasjoukkueen tasuri-% kotona
x5	kotijoukkueen tasuri-% kotona	x34	vierasjoukkueen tappio-% kotona
x6	kotijoukkueen tappio-% kotona	x35	vierasjoukkueen tehdyt maalit / ottelu kotona
x7	kotijoukkueen tehdyt maalit / ottelu kotona	x36	vierasjoukkueen päästetyt maalit / ottelu kotona
x8	kotijoukkueen päästetyt maalit / ottelu kotona	x37	vierasjoukkue pisteitä / ottelu kotona
x9	kotijoukkue pisteitä / ottelu kotona	x38	vierasjoukkue otteluita vieraissa
x10	kotijoukkue otteluita vieraissa	x39	vierasjoukkueen voitto-% vieraissa
x11	kotijoukkueen voitto-% vieraissa	x40	vierasjoukkueen tasuri-% vieraissa
x12	kotijoukkueen tasuri-% vieraissa	x41	vierasjoukkueen tappio-% vieraissa
x13	kotijoukkueen tappio-% vieraissa	x42	vierasjoukkueen tehdyt maalit / ottelu vieraissa
x14	kotijoukkueen tehdyt maalit / ottelu vieraissa	x43	vierasjoukkueen päästetyt maalit / ottelu vieraissa
x15	kotijoukkueen päästetyt maalit / ottelu vieraissa	x44	vierasjoukkue pisteitä / ottelu vieraissa
x16	kotijoukkue pisteitä / ottelu vieraissa	x45	vierasjoukkue otteluita yhteensä
x17	kotijoukkue otteluita yhteensä	x46	vierasjoukkueen voitto-% yhteensä
x18	kotijoukkueen voitto-% yhteensä	x47	vierasjoukkueen tasuri-% yhteensä
x19	kotijoukkueen tasuri-% yhteensä	x48	vierasjoukkueen tappio-% yhteensä
x20	kotijoukkueen tappio-% yhteensä	x49	vierasjoukkueen tehdyt maalit / ottelu yhteensä
x21	kotijoukkueen tehdyt maalit / ottelu yhteensä	x50	vierasjoukkueen päästetyt maalit / ottelu yhteensä
x22	kotijoukkueen päästetyt maalit / ottelu yhteensä	x51	vierasjoukkue pisteitä / ottelu yhteensä
x23	kotijoukkue pisteitä / ottelu yhteensä	x52	vierasjoukkue otteluita kuntopuntarissa
x24	kotijoukkue otteluita kuntopuntarissa	x53	vierasjoukkueen voitto-% kuntopuntarissa
x25	kotijoukkueen voitto-% kuntopuntarissa	x54	vierasjoukkueen tasuri-% kuntopuntarissa
x26	kotijoukkueen tasuri-% kuntopuntarissa	x55	vierasjoukkueen tappio-% kuntopuntarissa
x27	kotijoukkueen tappio-% kuntopuntarissa	x56	vierasjoukkueen tehdyt maalit / ottelu kuntopuntarissa
x28	kotijoukkueen tehdyt maalit / ottelu kuntopuntarissa	x57	vierasjoukkueen päästetyt maalit / ottelu kuntopuntarissa
x29	kotijoukkueen päästetyt maalit / ottelu kuntopuntarissa	x58	vierasjoukkue pisteitä / ottelu kuntopuntarissa

Tietokanta sisältää ottelukohtaisesti tiedot 58 eri muuttujasta. Taulukon kaksi ensimmäistä muuttujaa kertovat ottelun lopputuloksen, x1 kotijoukkueen maalimäärän ja x2 vierasjoukkueen maalimäärän. Tietokannassa on kahdeksan muuttujaa (x3, x10, x17, x24, x31, x38, x45, x52), jotka ilmaisevat pelattujen otteluiden määriä. Loput 48 muuttujaa mittaavat koti- ja vierasjoukkueen menestystä ennen ottelua.

Tietokanta sisältää ennustettavan kauden lisäksi kahden edellisen kauden otteluita. Laskettaessa parametreja tietokantaan jätetään osa otteluista huomioimatta. Tietokannan sisältöä on havainnollistettu kuvassa 2.

[18] Kuntopuntarissa on huomioitu joukkueen neljä viimeisintä ottelua.

Kuva 2: Havainnollistus tietokannan sisältämistä otteluista.



Kauden alusta jätetään pois jokaisen joukkueen osalta kuusi ensimmäistä ottelua. Koska sarjakausien välillä joukkueiden tasossa tapahtuu muutoksia pelaajakauppojen takia, ei edellisen kauden tuloksia voida hyödyntää ensimmäisten kierrosten ennusteiden laskemisessa. Osa joukkueista on myös vaihtunut, koska vuosittain huonoimmat joukkueet joutuvat alempaan sarjaan ja sieltä nousee vastaavasti uusia joukkueita tilalle. Näiden uusien joukkueiden otteluina pitäisi käyttää joko niiden edellisen kauden otteluita alemmalla sarjatasolla tai pudonneiden joukkueiden otteluita ylimmällä sarjatasolla. Kumpikaan vaihtoehto ei ole hyvä, joten on parempi odottaa kuusi kierrosta ja seurata, miten uudet joukkueet ovat menestyneet. Lisäksi kuuden ottelun jälkeen joukkueen neljän ottelun kuntopuntari eroaa jo hiukan koko sarjan menestyksestä.

Myös kausien lopusta jätetään tallentamatta jokaiselta joukkueelta kaksi viimeistä ottelua. Tämä johtuu siitä, että usein kauden lopussa joukkueiden motivaatiotekijät saattavat vaihdella hyvinkin paljon. Kärki- ja häntäpään joukkueet yrittävät tosissaan viimeiseen otteluun asti, kun taas keskikastin joukkueet saattavat viimeisissä otteluissa ajaa sisään nuoria pelaajiaan, kokeilla uusia pelitaktiikoita tai keskittyä kenties esimerkiksi mahdollisiin cup-otteluihin. Tällöin viimeisten kierrosten poikkeavat tulokset saattaisivat vääristää mallille estimoitavia parametreja.

Otteluiden rajaamisella pois on sekä positiivisia että negatiivisia seurauksia. Mitä pienempi otos kauden keskivaiheilta otetaan, sitä laadukkaampaa on kerätty data ja mallin ennusteiden tarkkuus paranee. Toisaalta tarkoituksena on hyödyntää ennusteita vedonlyönnissä, joten pelaajan kannalta optimaalisinta olisi etsiä kannattavia pelikohteita mahdollisimman suuresta määrästä otteluita. Voidaan todeta, että pelaajan on tehtävä vaihtokauppa ennusteiden tarkkuuden ja ennustettavien otteluiden määrän välillä. Työssä esitetty rajaus on eräänlainen välimuoto eikä välttämättä johda optimaaliseen tuottoon. Erilaisten rajausstrategioiden tarkastelua ei kuitenkaan tässä työssä käsitellä.

4.5 Faktorianalyysi muuttujille

Koska ennustemallin parametrit estimoidaan kuvan 2 mukaisesti kahden edellisen kauden tuloksilla, on aineistona tästä eteenpäin 221 Veikkausliigan ottelua kausilta 2002-2003. Aloitetaan muuttujien tarkastelu suorittamalla faktorianalyysi kaikille 48 muuttujalle. Tässä faktorien lukumäärän kriteerinä käytetään ominaisarvokriteeriä eli faktorimalliin otetaan mukaan ne faktorit joiden ominaisarvo on yli yksi. Tulkinnallisuuden helpottamiseksi faktoriratkaisu on rotatoitu. Tässä esitetään lyhyesti saatu faktoriratkaisu ja faktorien tulkinta. Kokonaisuudessaan alkuperäisten muuttujien faktorilataukset on esitetty liitteessä C.

Taulukko 16: Yhteenveto 48 muuttujan faktoriratkaisusta Veikkausliigan 221 ottelun aineistossa.

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	12.23	25.47	25.47
2	12.03	25.07	50.54
3	3.72	7.76	58.29
4	3.50	7.28	65.58
5	2.42	5.04	70.62
6	2.10	4.38	75.00
7	1.62	3.38	78.38
8	1.55	3.22	81.60
9	1.43	2.97	84.57
10	1.25	2.59	87.16
11	1.04	2.16	89.32
12	1.03	2.14	91.47

Nähdään, että on olemassa 12 kriteerin täyttävää faktoria. Näihin on saatu tiivistettyä hieman yli 90 % alkuperäisten muuttujien yhteisvaihtelusta. Tutkimalla liitteessä C esitettyjä faktorilatauksia voidaan faktorit tulkita ja antaa niille nimet, jotka on esitetty taulukossa 17.

Taulukko 17: Faktoreiden nimet.

FAKTORI	TULKINTA
1	kotijoukkueen yleismenestys
2	vierasjoukkueen yleismenestys
3	vierasjoukkueen kotimenestys
4	kotijoukkueen kotimenestys
5	kotijoukkueen nykykunto
6	vierasjoukkueen nykykunto
7	kotijoukkue ei-tasuria vieraissa
8	vierasjoukkue ei-tasuria kotona
9	vierasjoukkue ei-tasuria vieraissa
10	kotijoukkue ei-tasuria kotona
11	vierasjoukkueen maalintekovoima
12	kotijoukkueen maalintekovoima

Faktorit 1 ja 2 mittaavat koti- ja vierasjoukkueen menestystä yleisesti koko sarjassa. Ne määräävät selkeästi joukkueiden perustason, jota muiden faktoreiden avulla voidaan hienosäätää. Tämä tulee ilmi myös taulukosta 16, sillä nämä kaksi faktoria selittävät yli 50 % muuttujien alkuperäisestä yhteisvaihtelusta. Faktorit 3 ja 4 mittaavat menestystä vielä erikseen kotiotteluissa. Tämä on perusteltua, sillä jotkut joukkueet ovat erittäin riippuvaisia kotiotteluistaan kun taas toiset joukkueet menestyvät yhtä hyvin sekä koti- että vieraskentillä. Faktorit 5 ja 6 mittaavat joukkueiden tämän hetkistä kuntotilannetta. On selvää, että viimeiset ottelut kertovat enemmän joukkueen tilasta kuin esimerkiksi kaksi kuukautta aiemmin pelatut ottelut. Faktoreiden 7-10 tulkinta on mielenkiintoinen, sillä ne kaikki voidaan luokitella ei-tasapeliä faktoreiksi. Mitä korkeampia pistemääriä joukkue siis näistä faktoreista saa, sitä epätodennäköisempänä tasapelin mahdollisuutta voidaan pitää. Nämä faktorit ovat mielenkiintoisia, koska usein joukkueiden tasoero voi olla selkeä ja vaikeutena on juuri tasapeliherkkyyden arviointi. Viimeiset kaksi faktoria mittaavat joukkueiden maalintekovoimaa.

Aiemmissa tutkimuksissa esimerkiksi Dixon ja Coles (1997) sekä Peel ja Thomas (1992) ovat listanneet osatekijöitä, jotka vaikuttavat koti- ja vierasjoukkueen odotettuihin maalimääriin ja tulisi huomioida ennustemallissa:

- kotiedun merkitys
- viimeisimmillä otteluilla on suurin painoarvo joukkueen tulevaa menestystä arvioitaessa
- joukkueen kyky tehdä maaleja ja joukkueen taipumus päästää maaleja
- erot joukkueiden vahvuudessa koti- ja vierasotteluissa
- otteluiden tuloksia analysoidessa on otettava huomioon myös vastustajajoukkueen taso.

Faktorianalyysista saatu tulos antaa tukea aiemmissa tutkimuksissa esitetyille osatekijöille. Alkuperäisten muuttujien taustalla näyttäisi olevan selkeitä piilomuuttujia, joiden avulla joukkueiden menestystä voidaan selittää. Seuraavaksi valitaan malliin sisällytettävät osatekijät ja päätetään, mitkä alkuperäiset muuttujat soveltuvat parhaiten näiden osatekijöiden mittaamiseen.

4.6 Muuttujien luokittelu

Edellisen luvun perusteella valitaan koti- ja vierasmaaleihin vaikuttavat osatekijät.

Kotijoukkueen maalimäärän odotusarvoon vaikuttavat:

- Kotijoukkueen menestys koko sarjassa (mitä parempi menestys, sitä korkeampi odotusarvo)
- Kotijoukkueen menestys kotona (mitä parempi menestys, sitä korkeampi odotusarvo)
- Kotijoukkueen kuntopuntari (mitä parempi menestys, sitä korkeampi odotusarvo)
- Vierasjoukkueen menestys koko sarjassa (mitä parempi menestys, sitä matalampi odotusarvo)
- Vierasjoukkueen menestys vieraissa (mitä parempi menestys, sitä matalampi odotusarvo)
- Vierasjoukkueen kuntopuntari (mitä parempi menestys, sitä matalampi odotusarvo)

Vierasjoukkueen maalimäärän odotusarvoon vaikuttavat:

- Kotijoukkueen menestys koko sarjassa (mitä parempi menestys, sitä matalampi odotusarvo)
- Kotijoukkueen menestys kotona (mitä parempi menestys, sitä matalampi odotusarvo)
- Kotijoukkueen kuntopuntari (mitä parempi menestys, sitä matalampi odotusarvo)
- Vierasjoukkueen menestys koko sarjassa (mitä parempi menestys, sitä korkeampi odotusarvo)
- Vierasjoukkueen menestys vieraissa (mitä parempi menestys, sitä korkeampi odotusarvo)
- Vierasjoukkueen kuntopuntari (mitä parempi menestys, sitä korkeampi odotusarvo).

Seuraavaksi valitaan alkuperäisistä muuttujista ne, jotka parhaiten mittaavat edellä lueteltuja osatekijöitä. Tämän luokittelun tulokset on esitetty taulukossa 18.

Taulukko 18: Maalilukuihin vaikuttavat osatekijät ja niiden mittaamiseen valitut alkuperäiset muuttujat.

K1 - kotijoukkueen menestys koko sarjassa	V1 - vierasjoukkueen menestys koko sarjassa
kotijoukkueen tehdyt maalit / ottelu koko sarjassa kotijoukkueen päästetyt maalit / ottelu koko sarjassa kotijoukkue pisteitä / ottelu koko sarjassa	vierasjoukkueen tehdyt maalit / ottelu koko sarjassa vierasjoukkueen päästetyt maalit / ottelu koko sarjassa vierasjoukkue pisteitä / ottelu koko sarjassa
K2 - kotijoukkueen menestys kotona	**V2 - vierasjoukkueen menestys vieraissa**
kotijoukkueen tehdyt maalit / ottelu kotona kotijoukkueen päästetyt maalit / ottelu kotona kotijoukkue pisteitä / ottelu kotona	vierasjoukkueen tehdyt maalit / ottelu vieraissa vierasjoukkueen päästetyt maalit / ottelu vieraissa vierasjoukkue pisteitä / ottelu vieraissa
K3 - kotijoukkueen kuntopuntari	**V3 - vierasjoukkueen kuntopuntari**
kotijoukkueen tehdyt maalit / ottelu kuntopuntarissa kotijoukkueen päästetyt maalit / ottelu kuntopuntarissa kotijoukkue pisteitä / ottelu kuntopuntarissa	vierasjoukkueen tehdyt maalit / ottelu kuntopuntarissa vierasjoukkueen päästetyt maalit / ottelu kuntopuntarissa vierasjoukkue pisteitä / ottelu kuntopuntarissa

Taulukosta nähdään, että jokaisen osatekijän mittaamiseen riittää kolme alkuperäistä muuttujaa. Tämä johtuu siitä, että osa alkuperäisistä muuttujista on suoraan määritettävissä muiden muuttujien avulla. Esimerkiksi voitto- ja tasapeliprosentti määrittävät suoraan kerättyjen pisteiden määrän[19].

Jokaisen osatekijän sisällä kolme muuttujaa mittaavat samaa asiaa, joten niistä voidaan pääkomponenttianalyysin avulla rakentaa uusi yhdistetty muuttuja. Suoritetaan pääkomponenttianalyysi erikseen jokaiselle osatekijälle (K1, K2, K3, V1, V2, V3). Ottamalla kaikista osatekijöistä mukaan ensimmäinen pääkomponentti saadaan selittävien muuttujien lukumääräksi mallissa kuusi kappaletta.

4.7 Pääkomponenttianalyysin tulokset

Pääkomponenttianalyysin avulla mallin selittävien muuttujien lukumäärä saatiin vähennettyä kuuteen kappaleeseen, rakentamalla alkuperäisistä muuttujista uusia yhdistettyjä muuttujia. Ensimmäinen osatekijä K1 mittaa kotijoukkueen menestystä koko sarjassa. Pääkomponenttianalyysin tuloksia sen osalta on esitetty taulukoissa 19 ja 20.

[19] Pistekeskiarvo = (voitto-%*3)+(tasuri-%*1).

Taulukko 19: Pääkomponenttianalyysin tulokset osatekijälle K1.

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2.39	79.67	79.67
2	0.54	17.86	97.52
3	0.07	2.48	100.00

Taulukko 20: Osatekijän K1 pääkomponenttimatriisi.

Muuttujat	Component 1
koti_tehdyt_yhteensä	0.846
koti_päästetyt_yhteensä	-0.848
koti_pisteitä_yhteensä	0.977

Osatekijän K1 ensimmäinen pääkomponentti sisältää lähes 80 % alkuperäisten kolmen muuttujan informaatiosta. Pääkomponenttimatriisista nähdään, että alkuperäiset muuttujat latautuvat ensimmäisessä pääkomponentissa luontevasti. Mitä enemmän kotijoukkue on sarjassa tehnyt maaleja tai kerännyt pisteitä, sitä parempana sen menestystä voidaan pitää. Vastaavasti mitä enemmän se on päästänyt maaleja, sitä heikommin joukkueen voidaan katsoa menestyneen.

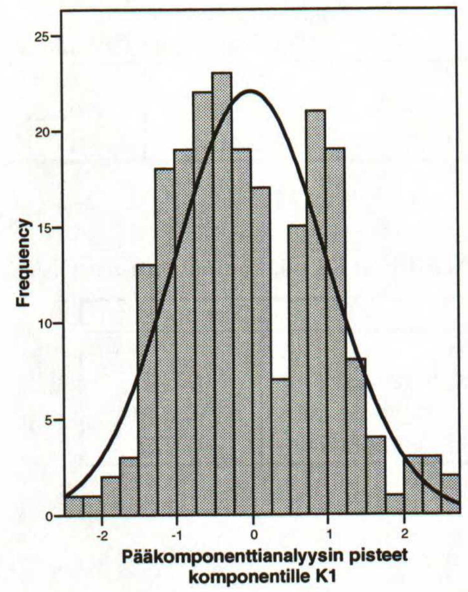
Verrattuna esimerkiksi pelkän sarjasijoituksen käyttämiseen selittävänä muuttujana ennustemallissa, osatekijöiden pääkomponenttipistemäärien käytöllä on kaksi selkeää etua. Ensinnäkin sarjataulun sijoitus mittaa järjestystä absoluuttisilla pistemäärillä, kun taas pääkomponentissa lasketaan suhteellista pistemäärää. Tällöin enemmän otteluita pelannut joukkue voi olla sarjataulukossa ylempänä, vaikka toinen joukkue on menestynyt suhteellisesti paremmin[20]. Toisekseen sarjataulukon sijaluku ei huomio tehtyjä ja päästettyjä maaleja, elleivät joukkueet ole samassa pistemäärässä. Sen sijaan pääkomponenttiin vaikuttavat kerätyn pistemäärän lisäksi maalimäärät. Tällöin pistemäärien valossa tasavahvojen joukkueiden välillä saattaa pääkomponenttipisteiden perusteella olla tasoero[21].

Koska pääkomponenttianalyysin avulla tuotettuja pistemääriä käytetään selittävinä muuttujina maalimäärien regressiomalleissa, tulisi niiden olla suunnilleen normaalijakautuneita. Osatekijän K1 pääkomponenttipisteiden jakauma on esitetty kuvassa 3.

[20] Esimerkiksi joukkue A kerännyt 10 ottelussa 20 pistettä ja joukkue B kerännyt 9 ottelussa 19 pistettä. Sarjataulukossa joukkue A on ykkösenä, mutta pääkomponentti asettaa ykköseksi joukkueen B sillä 19/9>20/10.

[21] Esimerkiksi joukkue A voittaa kaikki kauden 5 ensimmäistä ottelua osin tuurillakin 1-0. Joukkue B voittaa kaikki 5 otteluaan 3-0. Kuudennessa ottelussa joukkueet A ja B kohtaavat sarjasijoitukseltaan tasavahvoina, mutta pääkomponentti pitää joukkuetta B suosikkina huomattavasti paremman maalieronsa takia.

Kuva 3: Histogrammi osatekijän K1 pääkomponenttipistemäärille.



Kuvasta nähdään, että pistemäärä noudattaa välttävästi standardoitua normaalijakaumaa. Muiden osatekijöiden kohdalla tulokset ovat hyvin samanlaisia.

4.8 Pääkomponenttien hyödyntämien regressiomalleissa

Osatekijöille laskettuja pääkomponenttipisteitä voidaan hyödyntää regressiomalleissa selittävinä muuttujina maalilukujen odotusarvon mallintamiseksi. Vaikka osatekijät ja pääkomponenttien pistemäärät muodostetaan jokaiselle sarjalle samalla tavalla, on regressiomallit räätälöitävä sarjakohtaisesti. Tämä johtuu siitä, että eri sarjat ovat hyvin eri luonteisia. Esimerkiksi Italian Serie A:ssa pelataan yleensä enemmän tasapelejä kuin muissa sarjoissa. Myös kotiedun merkitys vaihtelee sarjoittain. Erään nyrkkisäännön mukaan kotietu on sitä suurempi, mitä etelämpänä euroopassa ollaan. Kolmanneksi sarjojen välillä on huomattavia eroja keskimääräisten maalimäärien suhteen. Useissa maissa otteluiden maalikeskiarvo on lähellä kolmea, mutta esimerkiksi Ranskassa vain hieman yli kaksi. Nämä kaikki tekijät vaikuttavat mallin parametreihin, ja siksi sarjakohtainen yksilöinti on välttämätöntä.

Käytännön perusteella voidaan tehdä oletuksia osatekijöiden ja maalimäärien korrelaatioista. Koska hyvä kotijoukkue tekee keskimäärin enemmän maaleja kuin huono, tulisi osatekijöiden K1, K2 ja K3 korreloida positiivisesti kotimaalien kanssa. Vastaavasti hyvä vierasjoukkue päästää keskimäärin vähemmän maaleja kotijoukkueelle kuin huono, joten kaikkien osatekijöiden V1, V2 ja V3 tulisi korreloida negatiivisesti kotimaalien kanssa. Vierasmaalien osalta tilanne on päinvastainen. Taulukossa 21 on esitetty osatekijöiden suorat korrelaatiot maalilukuihin Veikkausliigassa.

Taulukko 21: Osatekijöiden korrelaatiot Veikkausliigan koti- ja vierasmaalilukujen kanssa.

Osatekijä	kotimaalit	vierasmaalit
K1	0.193 **	-0.187 **
K2	0.150 *	-0.084
K3	0.171 *	-0.083
V1	-0.240 **	0.028
V2	-0.266 **	0.063
V3	-0.221 **	0.028

* = merkitsevä 0.05 riskitasolla (2-tailed)
** = merkitsevä 0.01 riskitasolla (2-tailed)

Kaikki suorat korrelaatiot ovat etumerkiltään odotetun kaltaisia. Kotimaalien kohdalla kaikki ovat myös tilastollisesti merkitseviä vähintään 0.05 riskitasolla, mutta vierasmaaleissa vain yksi osatekijä on tilastollisesti merkitsevä.

Suurin vaikutus maalimääriin on koti- ja vierasjoukkueen menestyksellä koko sarjassa. Nämä osatekijät K1 ja V1 määräävät tietynlaisen perustason joukkueille, jota sitten hienosäädetään muiden osatekijöiden avulla. Koska pääkomponenttianalyysi tehtiin erikseen jokaiselle osatekijälle (K1, K2, K3, V1, V2, V3), ne korreloivat myös keskenään. Tämän seurauksena osatekijöiden suorat korrelaatiot maalilukujen kanssa eivät anna oikeaa kuvaa maalilukujen ja osatekijöiden korrelaatiosta. Harhaton kuva osatekijöiden K2, K3, V2 ja V3 korrelaatiosta maalilukuihin saadaan tarkastelemalla niiden osittaiskorrelaatioita maalilukuihin. Tämä tarkoittaa osatekijöiden ja maalilukujen korrelaatioiden tarkastelua ehdolla, että tekijöiden K1 ja V1 vaikutus on vakioitu. (Laininen 2000, 106)

Osatekijöiden K2, K3, V2 ja V3 osittaiskorrelaatiot maalilukuihin on esitetty taulukossa 22.

Taulukko 22: Osittaiskorrelaatiot maalilukuihin kun tärkeimpien osatekijöiden K1 ja V1 vaikutus vakioitu.

Osatekijä	kotimaalit	vierasmaalit
K2	-0.008	0.132
K3	0.040	0.082
V2	-0.122	0.078
V3	-0.076	0.015
Osatekijöiden K1 ja V1 vaikutus vakioitu.		

Taulukosta nähdään, että yksikään osittaiskorrelaatio ei ole tilastollisesti merkitsevä. Kotimaalien kohdalla osatekijät K3, V2 ja V3 ovat etumerkiltään odotetun kaltaiset. Vierasmaalien kohdalla osatekijät V2 ja V3 saavat odotetut etumerkit. Näiden tulosten perusteella testataan taulukoissa 23 ja 24 esitettyjä ennustemalleja.

Taulukko 23: Testattavat mallit λ_h ennustamiseksi Veikkausliigassa. (x=osatekijä mukana mallissa)

				OSATEKIJÄT			
MALLI	vakio	K1	K2	K3	V1	V2	V3
1	x	x			x		
2	x	x		x	x		
3	x	x			x	x	
4	x	x			x		x
5	x	x		x	x	x	
6	x	x		x	x		x
7	x	x			x	x	x
8	x	x		x	x	x	x

Taulukko 24: Testattavat mallit λ_a ennustamiseksi Veikkausliigassa. (x=osatekijä mukana mallissa)

				OSATEKIJÄT			
MALLI	vakio	K1	K2	K3	V1	V2	V3
1	x	x			x		
2	x	x			x	x	
3	x	x			x		x
4	x	x			x	x	x

Käyttämällä tietokannan historiadataa voidaan parametrit estimoida eri malleille ja tutkia niiden kyvykkyyttä. Veikkausliigassa parhaat tulokset kotimaaleille λ_h saatiin mallilla 2 ja vierasmaaleille λ_a mallilla 1. Tällöin maalimäärien odotusarvot Veikkausliigassa ovat

$$\ln\lambda_{hi} = \beta_{h0} + \beta_{h1}k_{1i} + \beta_{h2}k_{3i} + \beta_{h3}v_{1i}$$

$$\ln\lambda_{ai} = \beta_{a0} + \beta_{a1}k_{1i} + \beta_{a2}v_{1i},$$

missä muuttujat k_1, v_1 ja k_3 ovat osatekijöiden K1, V1 ja V3 pääkomponenttipistemäärät ottelussa i.

Kotietu saadaan laskettua odotusarvojen erotuksena λ_h-λ_a. Tässä luvussa esitetyllä tavalla voidaan etsiä kaikille muillekin sarjoille sellaiset osatekijöiden yhdistelmät, jotka parhaiten soveltuvat maalimäärien odotusarvojen λ_h ja λ_a ennustamiseen. Kun parhaat osatekijöiden yhdistelmät on saatu selville, siirrytään tutkimaan, mikä esitetyistä regressiomalleista on paras.

4.9 Maali-mallit

Tässä luvussa tutkitaan pienellä aineistolla, mikä esitetyistä regressiomalleista on toimivin. Tutkittavat mallit ovat Poisson-regressio, NegBin-regressio ja BVDIP-regressio. Testiaineistona käytetään 85 Veikkausliigan ottelua ajalta 5.7-22.9.2004. Vaikka aineisto on suppea, käsittää se lähes 50% kauden 2004 otteluista. Aineiston perusteella ei ole vielä tarkoitus selvittää mallien kykyä voitolliseen vedonlyöntiin vaan ainoastaan tutkia eroja mallien välillä. Luvussa 5 parhaan mallin toimivuutta käytännössä testataan laajemmalla aineistolla.

4.9.1 Poisson-malli

Ensimmäinen malli perustuu Poisson-regressiolle. Tehdyn oletuksen mukaan kotimaalit (h) ja vierasmaalit (a) ovat toisistaan riippumattomia ja noudattavat Poisson-jakaumaa. Tuloksen (h,a) pistetodennäköisyys ottelussa i saadaan laskettua

$$f_{Poisson}(h,a \mid \lambda_{hi},\lambda_{ai}) = \left(\frac{e^{-\lambda_{hi}}\lambda_{hi}^{h}}{h!}\right)\left(\frac{e^{-\lambda_{ai}}\lambda_{ai}^{a}}{a!}\right) = p_{ha}, \tag{4.4}$$

missä

$$\ln\lambda_{hi} = \beta_{h0} + \beta_{h1}k_{1i} + \beta_{h2}k_{3i} + \beta_{h3}v_{1i}$$

$$\ln\lambda_{ai} = \beta_{a0} + \beta_{a1}k_{1i} + \beta_{a2}v_{1i}.$$

Pistetodennäköisyyksien p_{ha} ($h=\{0,1,...,h_{max}$[22]$\}$ ja $a=\{0,1,...,a_{max}\}$) perusteella syntyy ottelun todennäköisyysmatriisi **P**, jonka arvoja on vielä korjattava aiemmin määritetyllä korjausmatriisilla $\mathbf{C}_{\text{Veikkausliiga}}$, jolloin ottelun tulosmatriisin **T** alkiot saadaan laskettua

$$\mathbf{T} = \mathbf{C}_{\text{Veikkausliiga}} \times \mathbf{P}. \tag{4.5}$$

Matriisin **T** avulla lasketaan kotivoiton todennäköisyys summaamalla tekijät t_{ha}, joille $h>a$. Tasapelin ja vierasvoiton todennäköisyydet saadaan vastaavasti ehdoilla $h=a$ ja $h<a$.

4.9.2 NegBin-malli

Toinen malli perustuu NegBin-regressiolle ja tässäkin tapauksessa muuttujat h ja a oletetaan riippumattomiksi. Muuttujien ylihajonnan ilmaisee parametri α, jolloin tuloksen (h,a) pistetodennäköisyys ottelussa i saadaan laskettua

$$f_{\text{NegBin}}(h,a \mid \lambda_{hi},\lambda_{ai},\alpha_h,\alpha_a) = \left(\frac{\Gamma(\varphi_h + h)}{\Gamma(\varphi_h)h!u_i^{\varphi_h}(1-u_i)^h}\right)\left(\frac{\Gamma(\varphi_a + a)}{\Gamma(\varphi_a)a!v_i^{\varphi_a}(1-v_i)^a}\right) = p_{ha}, \tag{4.6}$$

missä

$\Gamma(.)$ = Gamma funktio

α_h = kotimaalien ylihajontaparametri

φ_h = $1/\alpha_h$

u_i = $\varphi_h / (\varphi_h + \lambda_{hi})$

α_a = vierasmaalien ylihajontaparametri

φ_a = $1/\alpha_a$

v_i = $\varphi_a / (\varphi_a + \lambda_{ai})$

$\ln\lambda_{hi} = \beta_{h0} + \beta_{h1}k_{1i} + \beta_{h2}k_{3i} + \beta_{h3}v_{1i}$

$\ln\lambda_{ai} = \beta_{a0} + \beta_{a1}k_{1i} + \beta_{a2}v_{1i}$.

---

[22] Jalkapallossa riittää kun lasketaan pistetodennäköisyydet siten, että $h_{max} = a_{max} = 7$. Koska jääkiekossa maalimäärän odutusarvo on korkeampi, riittää kun jääkiekossa lasketaan pistetodennäköisyydet siten, että $h_{max} = a_{max} = 10$.

Kuten Poisson-regression tapauksessa, syntyy ottelun todennäköisyysmatriisi **P** pistetodennäköisyyksien perusteella ja tulosmatriisin **T** alkiot saadaan matriisien **P** ja $\mathbf{C}_{\text{Veikkausliiga}}$ avulla

$$\mathbf{T} = \mathbf{C}_{\text{Veikkausliiga}} \times \mathbf{P}. \tag{4.7}$$

4.9.3 BVDIP-malli

Kolmas malli perustuu BVDIP-regressiolle. Tuloksen (h,a) pistetodennäköisyys ottelussa i saadaan

$$f_{\text{BVDIP}}(h, a \mid \lambda_{hi}, \lambda_{ai}, \lambda_{3i}) = e^{-(\lambda_{hi}+\lambda_{ai}+\lambda_{3i})} \frac{\lambda_{hi}^{h}}{h!} \frac{\lambda_{ai}^{a}}{a!} \sum_{j=0}^{\min(h,a)} \binom{h}{j}\binom{a}{j} j! \left(\frac{\lambda_{3i}}{\lambda_{hi}\lambda_{ai}}\right)^{j}, \tag{4.8}$$

missä

$$\ln\lambda_{hi} = \beta_{h0} + \beta_{h1}k_{1i} + \beta_{h2}k_{3i} + \beta_{h3}v_{1i}$$

$$\ln\lambda_{ai} = \beta_{a0} + \beta_{a1}k_{1i} + \beta_{a2}v_{1i}$$

$$\ln\lambda_{3i} = L.$$

Termi L on sarjakohtaisesti estimoitava vakiotermi, joka korjaa muuttujien λ_h ja λ_a mahdollista riippuvuutta[23]. Mallissa lisätään tasapelin todennäköisyyttä kasvattamalla diagonaalin solujen todennäköisyyksiä. Tällöin tulosmatriisi **T** syntyy pistetodennäköisyyksistä seuraavasti

$$T_{ha}(h,a) = \begin{cases} (1-p) f_{\text{BVDIP}}(h, a \mid \lambda_{hi}, \lambda_{ai}, \lambda_{3i}), & h \neq a \\ (1-p) f_{\text{BVDIP}}(h, a \mid \lambda_{hi}, \lambda_{ai}, \lambda_{3i}) + p f_{\text{Disc}(3)}(h \mid \theta), & h = a. \end{cases} \tag{4.9}$$

Diagonaalin todennäköisyyttä on tässä työssä jalkapallo-otteluille kasvatettu jakaumalla Disc(3), jolloin tasapelien todennäköisyyksiä on nostettu tuloksille 0-0, 1-1, 2-2 ja 3-3. Jääkiekossa maalimäärät ovat suurempia ja sen osalta diagonaalille on sovitettu Poisson-jakauma. Diagonaalin jakauman parametrit estimoidaan aina sarjakohtaisesti, koska tasapeliherkkyys on eri sarjoissa

[23] Mikäli L = 0, niin funktio pelkistyy Poisson-malliksi. Aiemmissa tutkimuksissa riippuvuustermiä on yritetty mallintaa myös selittävien muuttujien avulla, mutta merkittävää parannusta pelkän vakiotermin käyttöön ei ole saavutettu. Myös tämän työn yhteydessä yritettiin sovittaa riippuvuustermille erilaisia komponenttirakenteita, mutta parhaaksi osoittautui pelkän vakiotermin käyttäminen.

erilainen. Veikkausliigassa tietokannasta estimoidut parametrit ovat: p=0.046, θ_0=0.858, θ_1=0.000, θ_2=0.142 ja θ_3=0.000. (Yksityiskohtainen esitys BVDIP-mallista ja sen estimoinnista Karlis & Ntzoufras 2003.)

4.10 Mallien vertailu

Mallien kyvykkyyden vertailu suoritetaan tutkimalla Veikkausliigan 85 ottelun ennustettuja maalimääriä ja todennäköisyyksiä. Ennusteet lasketaan jokaiselle ottelukierrokselle käyttäen hyväksi vain siihen asti pelattuja aiempia otteluita, jolloin saadaan realistinen kuva mallien toiminnasta. Vaikka tässä vaiheessa ainoa tarkoitus on tutkia mallien välisiä eroja, otetaan tarkasteluun myös Tip-Ex verkkopalvelusta otetut todennäköisyysarviot. Kyseinen palvelu laskee todennäköisyyden kaikkien suurimpien vedonvälittäjien keskiarvona, joten sen arviota voidaan pitää markkinoiden ennusteena ottelun todennäköisyyksistä.

Ottelukohtaisesti lasketut maalien odotusarvot ja todennäköisyydet yhdessä markkina-arvioiden kanssa on esitetty kokonaisuudessaan liitteessä A. Tässä tuloksia käsitellään ainoastaan summatasolla. Tutkitaan ensiksi maalimäärien tarkkuutta, josta tulokset on esitetty taulukossa 25.

Taulukko 25: Koti- ja vierasmaalien ennustetut ja toteutuneet määrät.

	kotimaalit	vierasmaalit	yhteensä
Poisson-malli	125.2	92.0	217.2
NegBin-malli	133.1	99.1	232.2
BVDIP-malli	133.7	100.8	234.5
Toteutunut	136	93	229

NegBin-mallin ja BVDIP-mallin kohdalla maalien yhteismäärä on kohtuullisen lähellä toteutunutta arvoa. Kuten nähdään, niin nämä mallit aliarvioivat kotimaalien määrää ja yliarvioivat vierasmaalien määrää. Jos laskettuja arvoja käytettäisiin suoraan sellaisenaan ottelun 1X2-todennäköisyyksien arviointiin, annettaisiin systemaattisesti vierasvoitoille liian suuria ja kotivoitoille liian pieniä todennäköisyyksiä. Poisson-malli aliarvioi kotimaaleja muita pahemmin ja vaikka vierasmaalien ennuste on kohtuullisen tarkka, jää kokonaismäärän ennuste varsin epätarkaksi. Kokonaisuudessaan mallien tarkkuutta voidaan pitää kohtuullisena.

Taulukossa 26 on esitetty mallien antamat keskimääräiset todennäköisyysarviot kotivoitolle, tasapelille ja vierasvoitolle. Toteutuneet arvot kertovat todellisen 1X2-jakauman Veikkausliigan 85 ottelun aineistossa.

Taulukko 26: Mallien keskimääräiset todennäköisyysarviot eri lopputuloksille[24].

	1	X	2
Poisson-malli	0.450	0.257	0.293
NegBin-malli	0.431	0.258	0.311
BVDIP-malli	0.442	0.272	0.285
Toteutunut	0.482	0.271	0.247

Mallien antamat keskimääräiset todennäköisyysarviot ovat hyvin samansuuntaisia. Suurin poikkeama on BVDIP-mallin kohdalla, koska se antaa tasapelille selkeästi suuremman todennäköisyyden kuin muut mallit. Verrattaessa mallien arvoja toteutuneeseen jakaumaan nähdään, että ennusteiden tarkkuudessa on huomattavasti parantamisen varaa. Tämä ei välttämättä johdu siitä, että mallit ovat huonoja vaan mahdollisesti poikkeuksellisesta aineistosta. Mallien parametrit estimoitiin tietokannassa olevien kahden edellisen kauden otteluiden perusteella, joissa 1X2-todennäköisyysjakauma on ollut 0.44-0.25-0.31 vastaavasti. Erityisesti koti- ja vierasvoittojen todennäköisyydet testiaineistossa ja historiadatassa eroavat merkittävästi.

Koska pallopeleihin liittyy aina satunnaisuutta, ei poikkeava aineisto ole suuri ongelma sillä tilanne on sama kaikille vedonlyönnin osapuolille. Menestyksen kannalta on oleellisinta, kenen arviot ovat olleet tarkimpia. Tätä tutkitaan tarkastelemalla 85 Veikkausliigan ottelun oikeiden lopputulosten todennäköisyysarvioita eri malleilla sekä markkinaennustetta (Tip-Ex). Luonnollisesti osapuoli, jonka todennäköisyysarviot kautta linjan toteutuneille tuloksille ovat suurimpia, on osannut arvioida kyseisten otteluiden todennäköisyydet parhaiten.

Kyvykkyydelle tarvitaan mitta, jolla voidaan valita paras malli. Tässä voidaan hyödyntää Kullback-Leibler -informaatiota, joka mittaa kahden jakauman etäisyyttä. Sen avulla voidaan määrittää malli, jonka tuottamat arviot ovat mahdollisimman lähellä todellista todennäköisyyttä. Koska todellisia oikeita todennäköisyyksiä ei tässä tapauksessa tunneta, käytetään kyvykkyyden estimaattorina log likelihood arvoa ℓ

[24] Keskimääräinen ennuste on laskettu jokaiselle merkille kaikkien otteluiden ennusteiden keskiarvona.

$$\ell = \sum_{i=1}^{m} \ln p_i , \qquad (4.10)$$

missä p_i tarkoittaa ottelun i (i=1,2,...,m) toteutuneen tuloksen todennäköisyysarviota p, joiden logaritmit[25] summataan yli kaikkien aineiston otteluiden. (Sakamoto et al. 1986, 45-46) Taulukossa 27 on koottuna mallien kyvykkyyttä mittaavia tunnuslukuja Veikkausliigan aineistossa.

Taulukko 27: Mallien kyvykkyyden tunnuslukuja.

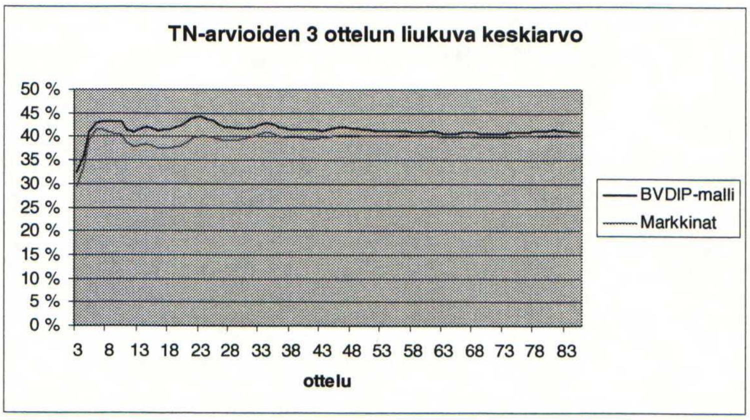
	ℓ	Σp_i	$\Sigma p_i/m$
Poisson-malli	-84.27	34.62	0.407
NegBin-malli	-83.39	34.08	0.401
BVDIP-malli	-82.30	34.70	0.408
Tip-Ex	-83.94	33.99	0.400

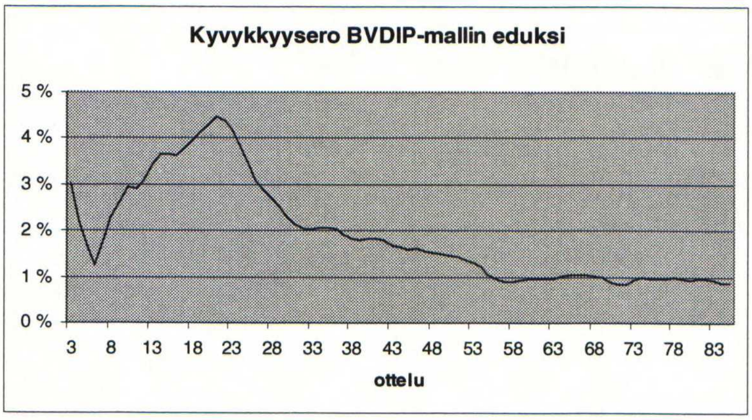
Jos kaikille toteutuneille tuloksille olisi annettu todennäköisyysarvioksi 100 %, saisi ℓ arvon 0. Mitä lähempänä arvo on nollaa, sitä tarkempina ennusteita voidaan pitää. Poisson-mallin ennusteet häviävät markkinaennusteelle, mutta NegBin-malli ja BVDIP-malli voittavat markkinaennusteen. Taulukon 27 tunnuslukujen perusteella BVDIP-malli erottuu parhaana mallina, joten valitaan BVDIP-regressio testatuista vaihtoehdoista tarkempaan jatkotarkasteluun.

Testatun aineiston koko on erittäin pieni eikä sen tulosten perusteella pidä tai voi tehdä luotettavia johtopäätöksiä parhaan mallin kyvykkyydestä suhteessa markkinoihin tai yksittäisiin vedonvälittäjiin. Tuloksia voidaan kuitenkin pitää rohkaisevina. Vaikka todellinen testaus suhteessa vedonvälittäjiin tehdään vasta seuraavassa luvussa, on BVDIP-mallin kyvykkyyttä suhteessa markkinaennusteisiin havainnollistettu kuvassa 4 Veikkausliigan osalta.

[25] Tässä on käytetty luonnollista logaritmia, mutta kantaluvun vaihtaminen ei vaikuta vertailun tulokseen.

Kuva 4: Markkinaennusteiden ja BVDIP-mallin ennusteiden vertailu Veikkausliigan aineistolla.





BVDIP-malli on kolmen ottelun liukuvalla keskiarvolla mitattuna koko kauden markkinoita kyvykkäämpi. Parhaimmillaan malli antaa yli 4 % tarkempia ennusteita kuin markkinakeskiarvo. Kauden loppua kohti mentäessä ero kutistuu. Tämä on luonnollista sillä kauden loppupuolella motivaatiotekijöiden merkitys korostuu eikä pelkkä tilastollinen analyysi tuota riittävän tarkkoja ennusteita. Tulos tukee aineiston rajauksen yhteydessä tehtyjä oletuksia. Mitä enemmän kausien alusta ja lopusta rajataan otteluita pois, sitä tarkempia ovat jäljelle jäävien otteluiden ennusteet. Tällöin kuitenkin myös aikajänne ennusteiden hyödyntämiseksi lyhenee.

Vaikka mallin tuottamat ennusteet olisivatkin hieman tarkempia kuin vedonvälittäjän ennusteet, ei tämä vielä takaa voitollista vedonlyöntiä, koska vedonvälittäjillä on etunaan deduktio. Seuraavassa luvussa seurataan minkälainen taloudellinen tulos eri sarjoissa saavutetaan, kun käytetään BVDIP-regressiota. Muuttujien λ_h ja λ_a mallintamiseen parhaiten sopivat osatekijöiden (K1, K2, K3, V1, V2, V3) rakenteet sekä regressiomallin kertoimet on estimoitu sarjakohtaisesti.

5. Tulokset

Tässä luvussa tutkitaan BVDIP-mallin toimivuutta käytännössä. Käytettävät kertoimet on otettu PAF rahapeliyhtiön pelipalvelusta. Aineistona ovat Suomen, Ruotsin, Norjan, Saksan, Ranskan, Espanjan, Englannin, Italian ja Brasilian pääsarjat jalkapallossa. Lisäksi mallin toimivuutta jääkiekkoon on testattu Ruotsin Elitserienin avulla. Suomen jääkiekon SM-liiga on jätetty tarkastelusta pois, koska sen pistelaskujärjestelmä muuttui kaudelle 2004-2005. Aineisto käsittää otteluita ajalta 07/2004-02/2005.

5.1 Pelikriteeri ja panostaminen

Panostettaessa yhteen otteluun eli pelattaessa ns. yksittäisveto, on Kellyn kaavan mukainen panostus optimaalinen (Hausch et al. 1994). Kellyn kaava on

$$b = \frac{pk - 1}{k - 1}, \tag{5.1}$$

missä b on panoksen osuus pelikassasta, p on voiton todennäköisyys ja k on kerroin. Oletus kaavan käytölle on, että pelaaja tietää eri lopputulosten oikeat todennäköisyydet. Tähän ei koskaan käytännössä päästä. Riskinhallinnan kannalta kaavan ehdottamaa panostusta on järkevää pienentää[26]. Tällöin pitkien tappioputkien aiheuttama riski pelaajan vararikosta pienenee. Tosin myös voittoputkien tapauksessa kassan kasvu on tällöin hitaampaa kuin käytettäessä täyttä Kelly panosta b.

Teoriassa jokainen kohde, jonka tuoton odotusarvo on yli yksi, on pelaajan kannalta voitollinen. Ennusteiden harhaisuudesta kuitenkin seuraa, että useissa aiemmissa tutkimuksissa (esim. Marttinen 2001, Dixon & Coles 1997) on havaittu pelaajan tuoton parantuvan kun kriteeriä on kiristetty[27]. Tässä työssä kriteeriä ei ole kiristetty, mutta sen sijaan panostusta on porrastettu. Menetelmää havainnollistetaan seuraavalla esimerkillä.

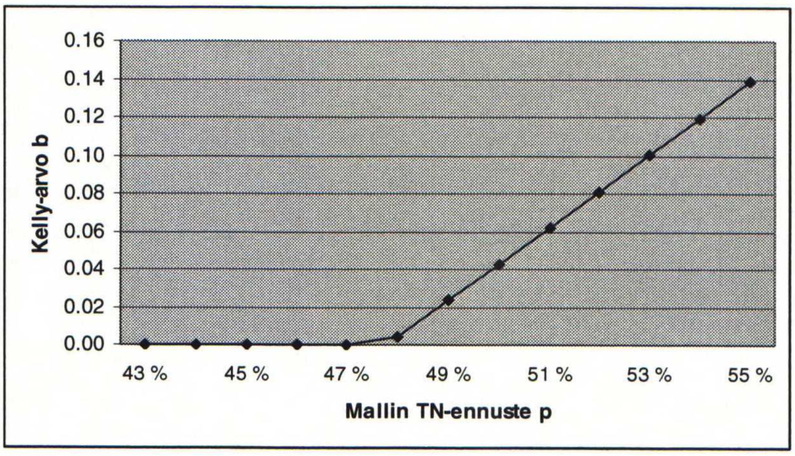
---

[26] Käytännössä yleisimpiä panostuksia ovat ½-Kelly ja ¼-Kelly. Esimerkiksi jälkimmäisessä otteluun sijoitetaan vain neljäsosa kaavan ehdottamasta panoksesta.

[27] Marttinen havaitsi tuottavimmaksi pelikriteeriksi odotusarvon 1,45. Dixon & Coles totesivat mallinsa pääsevän voitolliseen tulokseen, jos pelikriteerinä käytetään odotusarvoa 1,1.

Oletetaan, että vedonvälittäjän arvio kohteen oikealle tulokselle on 43 %, jolloin vedonvälittäjä tarjoaa kohteesta kerrointa 2.10[28]. Kertoimen ja mallin antaman todennäköisyysarvion perusteella laskettu Kelly-arvo b vaihtelee kuvan 5 osoittamalla tavalla.

Kuva 5: Mallin tn-arvion vaikutus kohteen Kelly-arvoon b.



Deduktion seurauksena kohteeseen kannattaa sijoittaa vasta, kun mallin antama arvio on yli 47.7 %. Jos Kelly-arvo on 0.1, niin mallin arvio on jo 10 % korkeampi kuin vedonvälittäjällä. Jos arvio on oikea, on mallin avulla löydetty todellinen huippukohde. Toisaalta jos arvio on väärä, sijoitetaan 10 % pelikassasta kohteeseen aivan liian suurella riskillä. Tätä riskiä tasataan työssä siten, että pienillä arvoilla ($0 < b \leq 0.05$) kohteeseen panostetaan arvon b mukainen määrä. Jos panossuositus $b > 0.05$, niin panostus kohteeseen on $^{b}/_{8}$. Tällöin ei jätetä hyödyntämättä mahdollisia huippukohteita, mutta virheellisen arvion tapauksessa ei kohteeseen panosteta suhteettoman suurta osaa pelikassasta.

5.2 Mallin menestys käytännössä

Aineisto käsittää yhteensä 1567 ottelua, joista BVDIP-mallin avulla löydettiin 625 pelikriteerin täyttävää kohdetta. Kohteisiin panostettiin yhteensä yli 15 000 euroa. Seurantajakson tulokset sarjakohtaisesti on esitetty taulukossa 28.

---

[28] Tässä vedonvälittäjän palautusprosentiksi on oletettu 90 %, jolloin (1/0.43)*0.9=2.10.

Taulukko 28: BVDIP-mallin ennusteilla laskettu pelikassan kehitys sarjoittain[29].

Maa	Sarja	Alkukassa	Panokset	Voitot	Loppukassa	PAL-%	Otteluita	Sijoitettu	osui	kerroinsumma	tas.PAL-%
Norja	Tippeligan	1 000 €	414 €	464 €	1 050 €	112.1 %	49	21	8	21.6	102.6 %
Suomi	Veikkausliiga	1 000 €	1 314 €	1 618 €	1 304 €	123.2 %	85	46	20	60.5	131.6 %
Italia	Serie A	1 000 €	1 401 €	1 578 €	1 177 €	112.6 %	129	63	18	64.7	102.6 %
Ruotsi	Allsvenskan	1 000 €	727 €	660 €	933 €	90.8 %	67	37	11	28.0	75.5 %
Brasilia	Serie A	1 000 €	1 513 €	2 108 €	1 594 €	139.3 %	175	50	18	59.7	119.3 %
Saksa	Bundesliiga	1 000 €	995 €	925 €	929 €	92.9 %	110	42	14	31.7	75.4 %
Espanja	Primera Div	1 000 €	1 269 €	919 €	650 €	72.4 %	149	65	12	34.8	53.5 %
Ranska	La Liga	1 000 €	1 495 €	1 170 €	675 €	78.2 %	167	74	18	50.3	68.0 %
Englanti	Valioliiga	1 000 €	3 676 €	4 190 €	1 514 €	114.0 %	185	101	31	97.8	96.8 %
Englanti	Champs	1 000 €	1 849 €	2 189 €	1 340 €	118.4 %	299	86	33	89.6	104.2 %
Ruotsi	Elitserien	1 000 €	981 €	1 084 €	1 102 €	110.4 %	152	40	14	31.6	78.9 %
	Yhteensä	11 000 €	15 635 €	16 904 €	12 269 €	108.1 %	1567	625	197	570.0	91.2 %

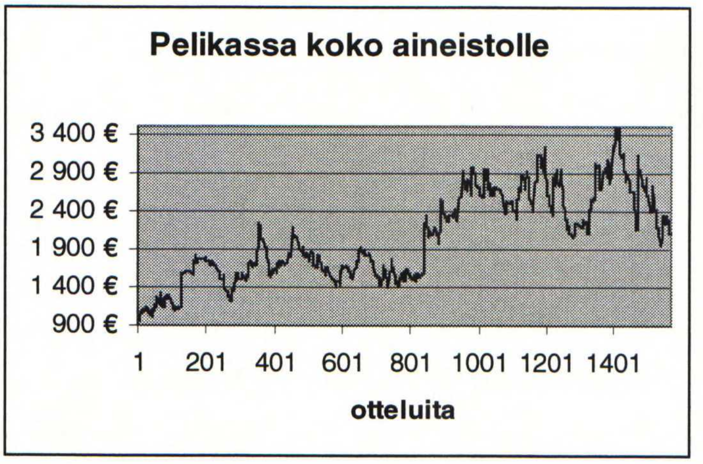
Kokonaisuudessaan seurantajakson tulos on hyvä. Mallin palautusprosentti on 108 % eli jokaisesta sijoitetusta eurosta on saatu voittoja 1.08 euroa. Tulos on erinomainen, kun muistetaan, että vedonvälittäjän laskennallinen palautusprosentti on 90 %. Mallin avulla kyettiin poistamaan vedonvälittäjän deduktio ja lisäksi tekemään voittoa. Tasapanostuksellakin laskien päästiin yli 91 % palautukseen.

Tuotolla mitattuna parhaiten menestyivät sijoitukset Brasilian Serie A:han ja Englannin Valioliigaan. Palautusprosentilla mitattuna hyviä tuloksia saatiin myös Suomen Veikkausliigassa ja Englannin Championship:ssa. Heikoiten sarjoista tuottivat sijoitukset Ranskan ja Espanjan sarjoihin, joissa palautusprosentti jäi alle 90 %. Näissä sarjoissa tulos on huonompi kuin mihin arpomalla olisi laskennallisesti päädytty.

Alkukassalle laskettu pääoman tuotto on 11.5 %. Todellisuudessa tuotto on huomattavasti suurempi, koska alkukassa on laskennallinen arvo eikä vaadi sitomaan koko summaa pelitoimintaan. Jos yhdistetään kaikki sarjat yhdeksi kokonaisuudeksi ja panostetaan aiemmin esitetyllä tavalla, kehittyy 1000 euron alkukassa kuvan 6 esittämällä tavalla.

[29] Sarake PAL-% tarkoittaa esitetyn Kelly-panostuksen tuottamaa palautusprosenttia. Sarake tas.PAL-% tarkoittaa palautusta, jos jokaiseen pelikriteerin täyttävään kohteeseen panostettaisiin aina vakio summa.

Kuva 6: Pelikassan kehitys tarkastelujaksolla koko aineistossa.



Pelikassa kasvaa alkuarvostaan tarkastelujaksolla 1112 euroa, joten alkupääoma on kuudessa kuukaudessa yli kaksinkertaistettu. Parhaimmillaan pelikassan arvo on tarkastelujakson aikana yli 3400 euroa. Vastaavat sarjakohtaiset pelikassan kehitystä kuvaavat graafit on esitetty liitteessä B.

Seurantajakso on varsin lyhyt ja arvioitujen kohteiden määrä kohtalaisen pieni, joten luotettavien johtopäätösten tekeminen on hankalaa. Tuloksia voidaan kuitenkin pitää rohkaisevina. Mallin tuottamaa informaatiota voidaan pitää hyvänä lähtökohtana voitolliselle pelaamiselle. Kun pelikohteet on laitettu tilastollisella analyysillä kannattavuuden perusteella järjestykseen, voidaan parhaiden kohteiden osalta suorittaa tarkempaa kvalitatiivista analyysiä. Tämän tuloksena pelaaja kykenee löytämään ne kohteet, joiden avulla voitollinen pelaaminen on mahdollista pitkällä aikavälillä. Työssä kaikki vedot pelattiin PAF:lle. Todellisuudessa jo muutaman vedonvälittäjän kertoimia vertailemalla tai pelaamalla pelit vedonlyöntipörssiin voitaisiin pelikohteille saada korkeampia kertoimia, jolloin mallin tuotto paranee entisestään.

6. Yhteenveto ja johtopäätökset

Tutkielman teoriaosan tavoitteena oli luoda yleiskäsitys vedonlyönnin peruskäsitteistä sekä aiemmista tutkimuksista. Lisäksi käytiin lyhyesti läpi työssä käytettyjä tilastollisia menetelmiä.

Työn empiriaosassa suoritettiin aluksi tilastollista testaamista, jotta voitiin tutkia maalilukuihin liitettyjen jakaumaoletusten mielekkyyttä. Tämän jälkeen esiteltiin työn tietokannan sisältämät muuttujat, joita analysoitiin faktorianalyysin avulla. Saadun faktoriratkaisun ja aiempien tutkimustulosten perusteella voitiin todeta muuttujien muodostavan luontevia osatekijöiden kokonaisuuksia, jotka vaikuttavat kotimaalien ja vierasmaalien lukumääriin. Näiden tulosten perusteella nimettiin kuusi osatekijää, joita käytettiin työssä maalimäärien ennustamiseen regressiomalleissa. Jokaiselle osatekijälle suoritettiin erikseen pääkomponenttianalyysi, jolla osatekijöiden mittaamiseen käytetyistä alkuperäisistä muuttujista rakennettiin uusia yhdistelmämuuttujia. Näiden pääkomponenttien pistemääriä käytettiin selittävinä muuttujina regressiomalleissa. Luvun 4 lopuksi testattiin kolmea regressiomallia ja valittiin parhaaksi BVDIP-malli. Valitun mallin toimivuutta käytännössä tutkittiin luvussa 5, käyttäen PAF:in tarjoamia 1X2-kertoimia.

Mallin avulla saavutetut tulokset olivat hyviä. Sijoitetuille panoksille saatiin palautusprosentiksi 108 %. Sidotulle pääomalle saatu tuotto todettiin huomattavasti suuremmaksi, sillä alkukassa on laskennallinen arvo eikä vaadi sitomaan koko rahasummaa pelaamiseen. Yhdistämällä kaikki ottelut yhdeksi kokonaisuudeksi saatiin alkukassan arvo yli kaksinkertaiseksi kuudessa kuukaudessa. Huomiota kiinnitettiin aineiston suhteellisen pieneen kokoon, jolloin luotettavien arvioiden tekeminen mallin kyvykkyydestä on vaikeaa. Kokonaisuudessaan työn tuloksia voidaan pitää rohkaisevina ja ne toimivat hyvänä pohjana kohteiden tarkemmalle kvalitatiiviselle analyysille.

Työssä esitetty malli on vain yksi ehdotus vedonlyönnin jatkuvasti kehittyvässä maailmassa. Aikaisempiin tutkimuksiin verrattuna työssä käytetty pääkomponenttianalyysi oli uusi menetelmä. Mallissa ei kuitenkaan huomioitu monia lopputulokseen vaikuttavia tekijöitä kuten loukkaantumisia, motivaatiota tai pelikieltoja. Näiden puutteiden takia voitolliseen tulokseen tähtäävän pelaajan on suoritettava kvalitatiivista analyysiä mallin löytämille parhaille kohteille.

Käytetyt menetelmät tarjoavat useita mahdollisuuksia jatkotutkimukselle ja mallin kehittämiselle. Loukkaantumiset ja pelikiellot olisi mahdollista rakentaa omaksi komponentiksi nykyisten lisäksi, mittaamalla niiden vaikutusta esimerkiksi poissaolijoiden tekemien maalien määrällä. Motivaation huomioimiseksi voitaisiin kehittää erilaisia menetelmiä, esimerkiksi joukkueen piste-ero putoamis- ja mitalitaisteluihin. Historiallisesti myös joukkueiden paikallisottelut ovat tasaisempia kuin pelkkä tilastollinen analyysi olettaa. Tällainen ulottuvuus voitaisiin lisätä malliin kohtuullisen helposti esimerkiksi dummy-muuttujan avulla. Myös työssä esitetty ongelma tietokannan otteluiden optimaalisesta rajaamisesta tarjoaisi mahdollisuuden tutkia mallin tuoton parantamisen edellytyksiä. Nämä kaikki vaatisivat kuitenkin tietokannalta dataa, jonka konstruointi takautuvasti on mahdotonta. Sen sijaan tulevaisuudessa näitä lisäyksiä olisi mahdollista tutkia ja siihen työssä esitetty malli antaa hyvän lähtökohdan.

Vedonlyönti on haastava harrastus tai ammatti. Vedonvälittäjien tietämys ja menetelmät kehittyvät jatkuvasti. Internetin vaikutuksesta muuttunut vedonlyöntiympäristö tarjoaa kuitenkin pelaajalle erinomaiset mahdollisuudet globaalisti markkinoiden ja kohteiden analysointiin ja hinnoitteluvirheiden hyväksikäyttöön. Nopean kehityksen seurauksena vanhat menetelmät ja teoriat eivät välttämättä toimi tulevaisuudessa ja koko ala elää siksi jatkuvassa muutostilassa. Tämä haasteellisuus ja jatkuva kehittymisen pakko motivoi kuitenkin innokkaat pelaajat jatkamaan ennusteiden laskemista kaudesta toiseen, vaikka jatkaminen tappioputken aikana saattaa tuntua turhauttavalta.

LÄHDELUETTELO

Audas, R. & Dobson, S. & Goddard, J. (1997). Team Performance and Managerial Change in the English Football League. Economic Affairs Vol. 17, No. 3, pp. 30-36.

Barnett, V. & Hilditch, S. (1993). The Effect of an Artificial Pitch Surface on Home Team Performance in Football. Journal of the Royal Statistical Society, Series A, 156, 39-50.

Bruinshoofd, A. & Weel, B. (2003). Manager to Go? Performance Dips Reconsidered with Evidence from Dutch Football. European Journal of Operational Research 148, 233–246.

Clarke, S. & Norman, J. (1995). Home Ground Advantage of Individual Clubs in English Soccer. The Statistician, 1995, 44, No 4, p509-521.

Crowder, M. & Dixon, M. & Ledford, A. & Robinson, M. (2002). Dynamic Modelling and Prediction of English Football League Matches for Betting. The Statistician 51:2. 157-168.

Dixon, M. & Coles, S. (1997). Modelling Association Football Scores and Inefficiencies in the Football Betting Market. Applied Statistics, 46(2), 265-280.

Dixon, M. & Robinson, M. (1998). A Birth Process Model for Association Football Matches. The Statistician 47:3. 523-538.

Dobson, S. & Goddard, J. (2003). Persistence in Sequences of Football Match Results: A Monte Carlo Analysis. European Journal of Operational Research, 148, 247–256.

Epstein, R. (1967). The Theory of Gambling and Statistical Logic. Academic Press Inc, New York.

Ernst & Young (2000). Winners and Losers - The Future of Online Betting.

Goddard, J. & Asimakopoulos, I. (2003). Modelling Football Match Results and the Efficiency of Fixed-Odds Betting. The University of Wales Swansea, Department of Economics.

Greene, W. (2000). Econometric Analysis, Fourth Edition, Prentice-Hall.

Haas, D. (2003). Technical Efficiency in the Major League Soccer. Journal of Sports Economics, Vol. 4, No. 3, 203-215. SAGE Publications.

Harman, H. (1976). Modern Factor Analysis. 3rd ed. Chicago: University of Chicago Press.

Hausch, D. & Lo, V. & Ziemba, W. (1994). Efficiency of Racetrack Betting Markets. Academic Press.

Hill, I. (1974). Association Football and Statistical Interference. Applied Statistics, 23(2), 203-208.

Hirotsu, N. & Wright, M. (2003). An Evaluation of Characteristics of Teams in Association Football by Using a Markov Process Model. The Statistician, vol. 52/4, pp 591-602.

Jackson, D. (1994). Index Betting on Sports. The Statistician 43:2. 309-315.

Juva, I. (2003). Jalkapallo-ottelun stokastinen mallintaminen. Sovelletun matematiikan harjoitustyö. Teknillinen korkeakoulu, Espoo.

Karlis, D. & Ntzoufras, J. (2003). Analysis of Sports Data Using Bivariate Poisson Models. The Statistician, 52, 381-393.

Koning, R. (2000). Balance in Competition in Dutch Soccer. The Statistician, vol. 49/3, pp. 419-431.

Koning, R. & Koolhaas, M. & Renes, G. & Ridder, G. (2003). A Simulation Model for Football Championships. European Journal of Operational Research 148, 268–276.

Kuonen, D. (1996). Modelling the Success of Soccer Teams in European Championships. Technical Report 96.1. Swiss Federal Institute of Technology, Lausanne.

Laininen, P. (2000). Tilastollisen analyysin perusteet., kolmas painos. Otatieto/Oy Yliopistokustannus, Helsinki.

Lee, A. (1999). Modelling Rugby League Data Via Bivariate Negative Binomial Regression. Australian & New Zealand Journal of Statistics, 1999, vol. 41, no. 2, pp. 141-152.

Lee, A. (1997). Modelling Scores in the Premier League: Is Manchester United Really the Best? Chance, vol. 10, 15-19.

Leskinen, E. (1987). Faktorianalyysi: konfirmatoristen faktorimallien teoria ja rakentaminen. Jyväskylän yliopiston tilastotieteen laitoksen julkaisuja 1.

Long, S. (1997). Regression Models for Categorical and Limited Dependent Variables. Advanced Quantitative Techniques in the Social Sciences, Volume 7. Sage Publications.

Maher, M. (1982). Modelling Association Football Scores. Statistica Neerlandica 36, 109–118.

Marttinen, N. (2001). Creating a Profitable Betting Strategy for Football by Using Statistical Modelling. Statistics Thesis, Trinity College Dublin.

McLachlan, G. & Krishnan, T. (1997). The EM Algorithm and Extensions. John Wiley & Sons, New York.

Mellin, I. (2004a). Tilastolliset monimuuttujamenetelmät (Mat-2.112) luentomoniste: Faktorianalyysi. Teknillinen korkeakoulu.

Mellin, I. (2004b). Tilastolliset monimuuttujamenetelmät (Mat-2.112) luentomoniste: Pääkomponenttianalyysi. Teknillinen korkeakoulu.

Moroney, M. (1956). Facts from Figures. Penguin Books Ltd, Baltimore.

Mustonen, S. (1995). Tilastolliset monimuuttujamenetelmät. Helsingin yliopisto, Tilastotieteen laitos.

National Gambling Impact Study Commission (1999). Final Report, Washington, D.C.

Neter, J. & Kutner, M. & Nachtsheim, C. & Wasserman, W. (1996). Applied Linear Statistical Models. 4th ed. Irwin, Homewood, IL.

Niemi, E. & Malmberg, J. (2003). Tilastotieteen yhdistelmäkurssi, luentomoniste. Helsingin kauppakorkeakoulu

Paakkulainen, H. (1995). Malli Pitkävetokerrointen osuvuuden testaamiseen. Tilastotieteen laitoksen Pro gradu-tutkielma, Joensuun yliopisto.

Peel, D. & Thomas, D. (1992). The Demand for Football: Some Evidence on Outcome Uncertainty. Empirical Economics 17:2. 323-331.

Raitanen, J. (1999). Jalkapallo-ottelun tilastollinen mallintaminen. Tilastotieteen pro gradu-tutkielma, Tampereen yliopisto.

Reep, C. & Benjamin, B. (1968). Skill and Chance in Association Football. Journal of the Royal Statistical Society A 131. 581-585.

Reep, C. & Pollard, R. & Benjamin, B. (1971). Skill and Chance in Ball Games. Journal of the Royal Statistical Society A 134. 623-629.

Ridder, G. & Cramer, J. & Hopstaken, P. (1994). Down to Ten: Estimating the Effect of a Red Card in Soccer. Journal of the American Statistical Association 89, 1124–1127.

Rue, H. & Salvesen, O. (2000). Predicting and Retrospective Analysis of Soccer Matches in a League. The Statistician, vol. 49, pp. 399–418.

Sakamoto, Y. & Ishiguro, M. & Kitagawa, G. (1986). Akaike Information Criterion Statistics. D. Reidel Publishing Company.

Vuoksenmaa, J. & Kuronen, A. & Nåls, J. (1999). Urheiluvedonlyönti - voittajan opas. Gummerus Kirjapaino Oy, Jyväskylä.

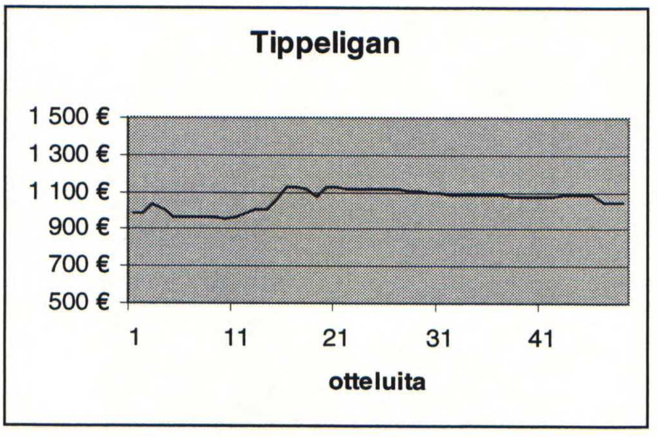
Liite A: Mallien ennusteet Veikkausliigan otteluihin

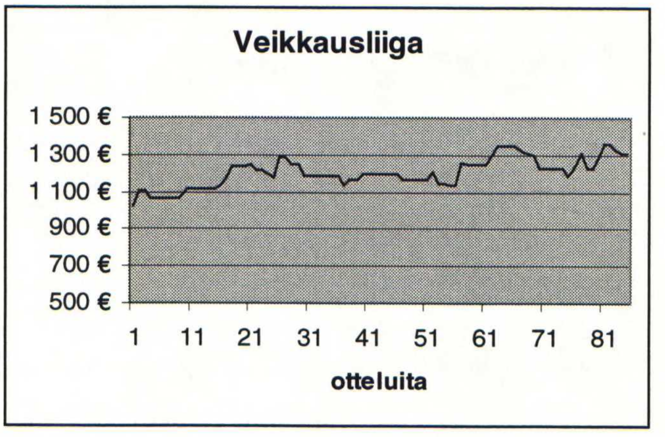
			MARKKINAT			Poisson-malli					TULOS
			todennäköisyydet			maalien odotusarvot		todennäköisyydet			
Päivä	Kotijoukkue	Vierasjoukkue	1	X	2	λ_h	λ_a	1	X	2	
05.07.2004	RoPS	Lahti	0.31	0.27	0.42	0.52	1.66	0.11	0.24	0.65	X (2-2)
05.07.2004	Jazz	Allianssi	0.27	0.27	0.46	1.23	0.96	0.41	0.30	0.29	X (1-1)
05.07.2004	Hämeenlinna	Inter	0.24	0.27	0.49	0.82	1.80	0.16	0.24	0.60	2 (3-4)
05.07.2004	HJK	TPS	0.56	0.25	0.19	1.30	1.04	0.42	0.29	0.29	1 (3-2)
05.07.2004	Haka	KooTeePee	0.60	0.24	0.16	1.81	0.69	0.64	0.23	0.13	1 (4-1)
07.07.2004	TP-47	Tampere_Utd	0.28	0.28	0.44	0.99	1.12	0.31	0.31	0.38	1 (1-0)
08.07.2004	AC Allianssi	FC Haka	0.37	0.28	0.34	0.83	1.60	0.19	0.26	0.55	X (0-0)
08.07.2004	FC Inter	FF Jaro	0.62	0.23	0.14	2.75	0.69	0.80	0.13	0.07	1 (5-1)
08.07.2004	FC Lahti	TPS	0.44	0.28	0.28	1.64	1.04	0.51	0.26	0.24	2 (1-2)
08.07.2004	RoPS	HJK	0.17	0.26	0.57	0.72	1.34	0.20	0.29	0.51	1 (2-0)
11.07.2004	FC KooTeePee	FC Jazz	0.49	0.27	0.24	1.52	1.37	0.40	0.26	0.34	X (1-1)
11.07.2004	TP-47	FC Hämeenlinna	0.54	0.26	0.20	2.60	0.95	0.72	0.16	0.12	1 (2-0)
11.07.2004	TPS	MyPa	0.42	0.28	0.30	1.61	1.03	0.50	0.26	0.24	1 (1-0)
14.07.2004	RoPS	FC Hämeenlinna	0.46	0.27	0.27	2.10	1.09	0.60	0.21	0.19	X (2-2)
15.07.2004	MyPa	FC KooTeePee	0.51	0.27	0.22	1.60	1.12	0.48	0.26	0.26	X (1-1)
15.07.2004	FF Jaro	Tampere United	0.27	0.28	0.45	0.70	1.37	0.18	0.29	0.53	2 (0-3)
15.07.2004	TP-47	FC Lahti	0.41	0.28	0.31	1.14	0.97	0.39	0.31	0.31	1 (3-1)
18.07.2004	FC Jazz	HJK	0.25	0.27	0.48	0.70	1.04	0.23	0.33	0.43	X (0-0)
18.07.2004	FC KooTeePee	FF Jaro	0.55	0.26	0.19	1.80	0.74	0.62	0.23	0.14	1 (2-0)
18.07.2004	TPS	TP-47	0.49	0.27	0.24	1.07	0.83	0.40	0.32	0.28	1 (3-2)
18.07.2004	FC Haka	RoPS	0.69	0.20	0.11	1.84	0.67	0.65	0.22	0.12	1 (1-0)
21.07.2004	Tampere United	FC Haka	0.35	0.29	0.36	1.08	0.92	0.38	0.32	0.30	2 (1-3)
21.07.2004	FC Hämeenlinna	MyPa	0.31	0.29	0.40	1.34	1.35	0.36	0.27	0.37	2 (1-2)
21.07.2004	FF Jaro	TP-47	0.38	0.29	0.33	1.10	1.41	0.29	0.28	0.44	1 (3-0)
22.07.2004	FC Lahti	FC Inter	0.36	0.28	0.36	1.04	1.40	0.27	0.28	0.45	1 (2-1)
22.07.2004	RoPS	AC Allianssi	0.27	0.28	0.45	1.03	1.24	0.30	0.29	0.41	1 (1-0)
24.07.2004	FF Jaro	FC Hämeenlinna	0.46	0.28	0.26	1.74	1.39	0.45	0.24	0.31	X (1-1)
24.07.2004	HJK	FC KooTeePee	0.62	0.24	0.14	1.96	0.95	0.60	0.22	0.17	1 (3-2)
25.07.2004	MyPa	FC Lahti	0.44	0.28	0.27	1.41	1.19	0.41	0.27	0.31	X (0-0)
25.07.2004	TP-47	FC Jazz	0.48	0.28	0.24	1.22	1.22	0.35	0.29	0.36	1 (3-2)
29.07.2004	Tampere_Utd	Jazz	0.57	0.25	0.18	1.71	0.89	0.56	0.25	0.19	1 (6-0)
29.07.2004	Inter	TP-47	0.59	0.24	0.17	2.15	0.67	0.71	0.19	0.10	1 (2-1)
29.07.2004	MyPa	Jaro	0.55	0.25	0.20	2.17	0.84	0.67	0.20	0.13	1 (2-1)
01.08.2004	FC Inter	MyPa	0.55	0.26	0.19	1.43	0.85	0.50	0.28	0.22	2 (0-1)
01.08.2004	AC Allianssi	TPS	0.49	0.27	0.24	1.50	1.08	0.46	0.27	0.27	X (0-0)
01.08.2004	FC Hämeenlinna	FC Jazz	0.37	0.28	0.35	1.46	1.24	0.42	0.27	0.32	X (2-2)
01.08.2004	RoPS	Tampere United	0.28	0.28	0.44	1.04	1.42	0.27	0.28	0.46	1 (2-0)
05.08.2004	FC Haka	FC Inter	0.51	0.27	0.22	2.10	0.68	0.70	0.19	0.10	1 (4-0)
05.08.2004	FF Jaro	AC Allianssi	0.29	0.28	0.43	0.93	1.24	0.27	0.30	0.43	2 (1-2)
05.08.2004	FC Lahti	Tampere United	0.33	0.29	0.38	1.34	1.08	0.42	0.28	0.30	1 (2-1)
05.08.2004	RoPS	TP-47	0.39	0.29	0.32	1.64	0.97	0.52	0.26	0.22	X (1-1)
08.08.2004	FC KooTeePee	FC Hämeenlinna	0.55	0.26	0.19	1.80	1.12	0.53	0.24	0.23	1 (1-0)
08.08.2004	TP-47	HJK	0.30	0.29	0.41	1.14	1.10	0.36	0.30	0.35	2 (1-2)
08.08.2004	TPS	RoPS	0.57	0.25	0.18	2.18	0.89	0.66	0.20	0.14	1 (1-0)
11.08.2004	Tampere United	TP-47	0.59	0.25	0.16	1.98	0.79	0.65	0.21	0.13	1 (3-2)
12.08.2004	HJK	FC Inter	0.51	0.27	0.22	2.13	0.70	0.70	0.19	0.10	2 (0-1)
12.08.2004	FC Lahti	FC Jazz	0.54	0.26	0.20	1.95	0.94	0.61	0.22	0.17	1 (4-0)
12.08.2004	MyPa	RoPS	0.60	0.25	0.16	2.11	0.82	0.67	0.20	0.13	2 (1-2)
12.08.2004	TPS	FC KooTeePee	0.53	0.27	0.21	1.61	0.82	0.56	0.26	0.19	1 (1-0)
20.08.2004	HJK	AC Allianssi	0.50	0.27	0.23	1.94	0.83	0.63	0.22	0.15	2 (0-1)
21.08.2004	TP-47	FC Haka	0.22	0.27	0.51	0.87	1.51	0.21	0.27	0.52	2 (1-4)
22.08.2004	FC Inter	FC KooTeePee	0.59	0.25	0.17	1.55	0.85	0.53	0.26	0.20	X (1-1)
22.08.2004	Tampere United	TPS	0.51	0.27	0.22	1.30	0.95	0.44	0.29	0.27	1 (3-1)
22.08.2004	FC Hämeenlinna	FC Lahti	0.24	0.27	0.49	0.99	1.38	0.26	0.28	0.46	2 (0-2)
26.08.2004	FC KooTeePee	Tampere United	0.30	0.28	0.42	1.03	0.96	0.35	0.32	0.33	2 (0-1)
26.08.2004	MyPa	TP-47	0.59	0.25	0.16	1.59	0.65	0.60	0.26	0.15	1 (2-0)
26.08.2004	TPS	FC Hämeenlinna	0.68	0.21	0.12	1.59	0.76	0.57	0.26	0.17	2 (2-3)
29.08.2004	AC Allianssi	FC Lahti	0.46	0.27	0.27	1.20	1.19	0.36	0.29	0.36	2 (1-2)
29.08.2004	FC Haka	FC Jazz	0.70	0.19	0.11	2.78	0.61	0.82	0.13	0.05	1 (1-0)
29.08.2004	FF Jaro	HJK	0.29	0.28	0.43	0.97	1.32	0.27	0.29	0.44	X (2-2)
29.08.2004	RoPS	FC Inter	0.31	0.28	0.41	1.28	1.35	0.34	0.27	0.38	1 (4-0)
29.08.2004	Tampere United	MyPa	0.46	0.27	0.27	1.45	1.35	0.39	0.26	0.35	X (1-1)
02.09.2004	HJK	MyPa	0.45	0.28	0.27	1.28	1.21	0.37	0.28	0.35	X (0-0)
02.09.2004	FC Lahti	FC Haka	0.26	0.28	0.46	1.07	1.48	0.27	0.27	0.46	2 (1-2)
02.09.2004	TP-47	FC KooTeePee	0.47	0.28	0.26	1.57	1.27	0.44	0.26	0.31	1 (1-0)
02.09.2004	FC Inter	FC Jazz	0.61	0.24	0.15	2.19	0.94	0.65	0.20	0.15	1 (3-0)
05.09.2004	AC Allianssi	FC KooTeePee	0.58	0.25	0.17	1.76	0.95	0.56	0.24	0.20	X (0-0)
09.09.2004	FC Jazz	RoPS	0.42	0.28	0.30	1.41	1.42	0.36	0.26	0.38	2 (0-1)
09.09.2004	FF Jaro	TPS	0.35	0.28	0.38	1.29	1.33	0.35	0.27	0.38	1 (1-0)
09.09.2004	FC Hämeenlinna	AC Allianssi	0.23	0.27	0.50	0.99	1.51	0.24	0.27	0.49	2 (2-3)
09.09.2004	MyPa	FC Inter	0.47	0.27	0.26	1.58	1.03	0.50	0.26	0.24	X (1-1)
10.09.2004	FC Haka	HJK	0.54	0.26	0.20	1.57	0.71	0.58	0.26	0.16	1 (3-2)
12.09.2004	RoPS	FC KooTeePee	0.49	0.27	0.24	1.31	0.95	0.44	0.29	0.27	2 (1-3)
12.09.2004	Tampere United	FF Jaro	0.63	0.23	0.14	1.82	0.96	0.57	0.24	0.19	1 (2-0)
13.09.2004	AC Allianssi	TP-47	0.64	0.23	0.13	2.41	0.73	0.75	0.17	0.09	1 (3-2)
13.09.2004	TPS	FC Haka	0.24	0.27	0.49	0.95	1.30	0.26	0.29	0.44	X (0-0)
14.09.2004	HJK	FC Lahti	0.50	0.27	0.23	1.33	1.16	0.40	0.28	0.32	X (2-2)
15.09.2004	Tampere United	FC Hämeenlinna	0.72	0.18	0.10	2.18	0.77	0.70	0.19	0.11	1 (1-0)
18.09.2004	FC Haka	MyPa	0.56	0.26	0.18	1.40	0.90	0.48	0.28	0.24	1 (4-3)
18.09.2004	FC Hämeenlinna	HJK	0.22	0.26	0.52	0.78	1.20	0.23	0.31	0.46	X (2-2)
18.09.2004	FC Inter	Tampere United	0.36	0.28	0.36	1.32	1.11	0.41	0.28	0.31	1 (1-0)
18.09.2004	FF Jaro	FC Lahti	0.36	0.28	0.36	0.93	1.48	0.23	0.27	0.50	1 (2-1)
18.09.2004	FC Jazz	TPS	0.31	0.28	0.41	1.34	1.28	0.37	0.27	0.35	X (1-1)
18.09.2004	FC KooTeePee	AC Allianssi	0.33	0.28	0.39	1.12	1.32	0.31	0.28	0.41	2 (0-2)
22.09.2004	AC Allianssi	Tampere United	0.40	0.29	0.31	1.42	0.98	0.46	0.28	0.26	X (0-0)

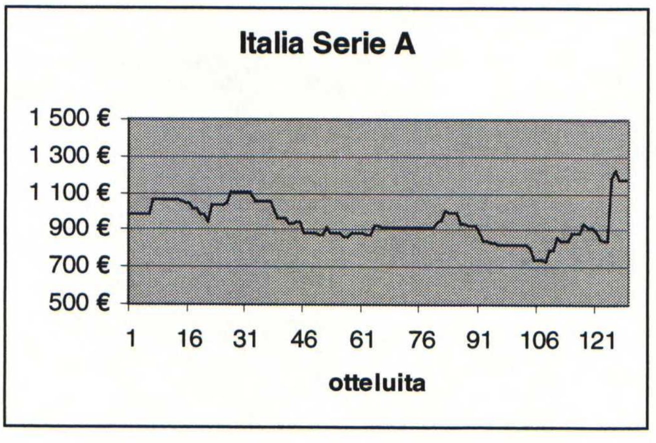
			MARKKINAT			NegBin-malli							TULOS
			todennäköisyydet			maalien odotusarvot		maalien varianssit		todennäköisyydet			
Päivä	Kotijoukkue	Vierasjoukkue	1	X	2	λ_h	λ_a	Var (λ_h)	Var (λ_a)	1	X	2	
05.07.2004	RoPS	Lahti	0.31	0.27	0.42	1.50	1.50	1.87	1.87	0.36	0.25	0.39	X (2-2)
05.07.2004	Jazz	Allianssi	0.27	0.27	0.46	2.20	2.20	2.99	2.99	0.39	0.19	0.42	X (1-1)
05.07.2004	Hämeenlinna	Inter	0.24	0.27	0.49	0.81	1.59	0.92	2.00	0.19	0.27	0.54	2 (3-4)
05.07.2004	HJK	TPS	0.56	0.25	0.19	1.48	0.97	1.83	1.13	0.45	0.28	0.27	1 (3-2)
05.07.2004	Haka	KooTeePee	0.60	0.24	0.16	1.99	0.63	2.64	0.70	0.64	0.23	0.13	1 (4-1)
07.07.2004	TP-47	Tampere_Utd	0.28	0.28	0.44	1.09	1.18	1.29	1.41	0.32	0.30	0.38	1 (1-0)
08.07.2004	AC Allianssi	FC Haka	0.37	0.28	0.34	0.90	1.42	1.04	1.75	0.23	0.28	0.48	X (0-0)
08.07.2004	FC Inter	FF Jaro	0.62	0.23	0.14	2.90	0.96	4.27	1.11	0.70	0.17	0.13	1 (5-1)
08.07.2004	FC Lahti	TPS	0.44	0.28	0.28	1.38	1.16	1.70	1.39	0.39	0.28	0.33	2 (1-2)
08.07.2004	RoPS	HJK	0.17	0.26	0.57	0.83	1.44	0.94	1.79	0.21	0.28	0.51	1 (2-0)
11.07.2004	FC KooTeePee	FC Jazz	0.49	0.27	0.24	1.42	1.05	1.74	1.24	0.42	0.28	0.30	X (1-1)
11.07.2004	TP-47	FC Hämeenlinna	0.54	0.26	0.20	2.45	1.13	3.43	1.33	0.61	0.20	0.19	1 (2-0)
11.07.2004	TPS	MyPa	0.42	0.28	0.30	1.74	1.01	2.23	1.18	0.50	0.26	0.24	1 (1-0)
14.07.2004	RoPS	FC Hämeenlinna	0.46	0.27	0.27	1.84	1.48	2.39	1.85	0.43	0.23	0.33	X (2-2)
15.07.2004	MyPa	FC KooTeePee	0.51	0.27	0.22	1.51	0.95	1.89	1.10	0.47	0.28	0.26	X (1-1)
15.07.2004	FF Jaro	Tampere United	0.27	0.28	0.45	1.01	1.85	1.17	2.41	0.20	0.24	0.56	2 (0-3)
15.07.2004	TP-47	FC Lahti	0.41	0.28	0.31	1.61	1.11	2.03	1.32	0.45	0.26	0.28	1 (3-1)
18.07.2004	FC Jazz	HJK	0.25	0.27	0.48	1.28	1.07	1.55	1.26	0.39	0.29	0.32	X (0-0)
18.07.2004	FC KooTeePee	FF Jaro	0.55	0.26	0.19	2.50	0.97	3.52	1.12	0.65	0.19	0.16	1 (2-0)
18.07.2004	TPS	TP-47	0.49	0.27	0.24	1.64	0.88	2.09	1.00	0.51	0.27	0.22	1 (3-2)
18.07.2004	FC Haka	RoPS	0.69	0.20	0.11	2.57	0.66	3.64	0.73	0.72	0.18	0.10	1 (1-0)
21.07.2004	Tampere United	FC Haka	0.35	0.29	0.36	1.03	1.01	1.20	1.18	0.33	0.32	0.35	2 (1-3)
21.07.2004	FC Hämeenlinna	MyPa	0.31	0.29	0.40	1.40	1.56	1.72	1.95	0.33	0.25	0.42	2 (1-2)
21.07.2004	FF Jaro	TP-47	0.38	0.29	0.33	1.09	1.39	1.28	1.71	0.28	0.28	0.44	1 (3-0)
22.07.2004	FC Lahti	FC Inter	0.36	0.28	0.36	0.93	1.15	1.08	1.37	0.28	0.31	0.41	1 (2-1)
22.07.2004	RoPS	AC Allianssi	0.27	0.28	0.45	1.22	1.53	1.47	1.91	0.29	0.26	0.44	1 (1-0)
24.07.2004	FF Jaro	FC Hämeenlinna	0.46	0.28	0.26	1.70	1.65	2.18	2.09	0.38	0.23	0.39	X (1-1)
24.07.2004	HJK	FC KooTeePee	0.62	0.24	0.14	1.85	1.00	2.42	1.17	0.53	0.25	0.23	1 (3-2)
25.07.2004	MyPa	FC Lahti	0.44	0.28	0.27	1.45	1.01	1.79	1.18	0.44	0.28	0.28	X (0-0)
25.07.2004	TP-47	FC Jazz	0.48	0.28	0.24	1.52	1.30	1.89	1.58	0.40	0.26	0.34	1 (3-2)
29.07.2004	Tampere_Utd	Jazz	0.57	0.25	0.18	1.70	0.78	2.18	0.89	0.55	0.26	0.19	1 (6-0)
29.07.2004	Inter	TP-47	0.59	0.24	0.17	2.21	0.98	3.01	1.13	0.60	0.22	0.18	1 (2-1)
29.07.2004	MyPa	Jaro	0.55	0.25	0.20	2.08	1.01	2.78	1.18	0.57	0.23	0.20	1 (2-1)
01.08.2004	FC Inter	MyPa	0.55	0.26	0.19	1.82	1.02	2.36	1.19	0.52	0.25	0.23	2 (0-1)
01.08.2004	AC Allianssi	TPS	0.49	0.27	0.24	1.35	1.27	1.65	1.54	0.36	0.27	0.36	X (0-0)
01.08.2004	FC Hämeenlinna	FC Jazz	0.37	0.28	0.35	1.48	1.51	1.84	1.88	0.35	0.25	0.39	X (2-2)
01.08.2004	RoPS	Tampere United	0.28	0.28	0.44	0.89	1.59	1.02	2.01	0.21	0.27	0.53	1 (2-0)
05.08.2004	FC Haka	FC Inter	0.51	0.27	0.22	1.65	0.71	2.10	0.79	0.55	0.27	0.18	1 (4-0)
05.08.2004	FF Jaro	AC Allianssi	0.29	0.28	0.43	1.37	1.79	1.67	2.32	0.29	0.24	0.47	2 (1-2)
05.08.2004	FC Lahti	Tampere United	0.33	0.29	0.38	1.12	1.12	1.32	1.33	0.33	0.30	0.36	1 (2-1)
05.08.2004	RoPS	TP-47	0.39	0.29	0.32	1.51	1.44	1.89	1.78	0.37	0.25	0.37	X (1-1)
08.08.2004	FC KooTeePee	FC Hämeenlinna	0.55	0.26	0.19	2.05	1.14	2.74	1.35	0.54	0.23	0.23	1 (1-0)
08.08.2004	TP-47	HJK	0.30	0.29	0.41	1.48	1.36	1.84	1.66	0.38	0.26	0.36	2 (1-2)
08.08.2004	TPS	RoPS	0.57	0.25	0.18	2.29	1.02	3.15	1.20	0.60	0.21	0.19	1 (1-0)
11.08.2004	Tampere United	TP-47	0.59	0.25	0.16	1.89	0.85	2.48	0.96	0.57	0.24	0.19	1 (3-2)
12.08.2004	HJK	FC Inter	0.51	0.27	0.22	1.76	1.10	2.27	1.29	0.49	0.25	0.26	2 (0-1)
12.08.2004	FC Lahti	FC Jazz	0.54	0.26	0.20	1.89	1.21	2.47	1.46	0.49	0.24	0.27	1 (4-0)
12.08.2004	MyPa	RoPS	0.60	0.25	0.16	2.16	1.04	2.93	1.22	0.58	0.22	0.20	2 (1-2)
12.08.2004	TPS	FC KooTeePee	0.53	0.27	0.21	1.84	0.98	2.39	1.14	0.53	0.25	0.22	1 (1-0)
20.08.2004	HJK	AC Allianssi	0.50	0.27	0.23	1.70	0.98	2.17	1.14	0.50	0.26	0.24	2 (0-1)
21.08.2004	TP-47	FC Haka	0.22	0.27	0.51	0.76	1.42	0.86	1.76	0.20	0.29	0.52	2 (1-4)
22.08.2004	FC Inter	FC KooTeePee	0.59	0.25	0.17	1.66	0.73	2.11	0.82	0.55	0.27	0.18	X (1-1)
22.08.2004	Tampere United	TPS	0.51	0.27	0.22	1.40	1.02	1.72	1.18	0.43	0.29	0.29	1 (3-1)
22.08.2004	FC Hämeenlinna	FC Lahti	0.24	0.27	0.49	1.15	1.42	1.36	1.74	0.29	0.27	0.43	2 (0-2)
26.08.2004	FC KooTeePee	Tampere United	0.30	0.28	0.42	1.40	1.29	1.73	1.56	0.38	0.27	0.36	2 (0-1)
26.08.2004	MyPa	TP-47	0.59	0.25	0.16	2.10	0.96	2.82	1.11	0.58	0.22	0.19	1 (2-0)
26.08.2004	TPS	FC Hämeenlinna	0.68	0.21	0.12	2.39	1.03	3.33	1.20	0.62	0.20	0.18	2 (2-3)
29.08.2004	AC Allianssi	FC Lahti	0.46	0.27	0.27	1.34	1.05	1.63	1.23	0.40	0.29	0.31	2 (1-2)
29.08.2004	FC Haka	FC Jazz	0.70	0.19	0.11	2.94	0.71	4.35	0.79	0.76	0.16	0.09	1 (1-0)
29.08.2004	FF Jaro	HJK	0.29	0.28	0.43	1.27	1.70	1.53	2.18	0.28	0.25	0.47	X (2-2)
29.08.2004	RoPS	FC Inter	0.31	0.28	0.41	1.19	1.43	1.42	1.76	0.30	0.27	0.43	1 (4-0)
29.08.2004	Tampere United	MyPa	0.46	0.27	0.27	1.28	1.08	1.54	1.27	0.38	0.29	0.33	X (1-1)
02.09.2004	HJK	MyPa	0.45	0.28	0.27	1.17	0.99	1.39	1.15	0.37	0.31	0.32	X (0-0)
02.09.2004	FC Lahti	FC Haka	0.26	0.28	0.46	0.89	1.48	1.02	1.84	0.22	0.28	0.50	2 (1-2)
02.09.2004	TP-47	FC KooTeePee	0.47	0.28	0.26	1.48	1.02	1.83	1.20	0.44	0.28	0.28	1 (1-0)
02.09.2004	FC Inter	FC Jazz	0.61	0.24	0.15	2.05	0.92	2.74	1.06	0.58	0.23	0.19	1 (3-0)
05.09.2004	AC Allianssi	FC KooTeePee	0.58	0.25	0.17	1.76	0.94	2.27	1.09	0.52	0.25	0.22	X (0-0)
09.09.2004	FC Jazz	RoPS	0.42	0.28	0.30	1.29	1.12	1.57	1.32	0.38	0.29	0.33	2 (0-1)
09.09.2004	FF Jaro	TPS	0.35	0.28	0.38	1.25	1.65	1.51	2.10	0.28	0.25	0.47	1 (1-0)
09.09.2004	FC Hämeenlinna	AC Allianssi	0.23	0.27	0.50	1.06	1.37	1.24	1.67	0.28	0.28	0.44	2 (2-3)
09.09.2004	MyPa	FC Inter	0.47	0.27	0.26	1.40	0.86	1.72	0.98	0.46	0.29	0.25	X (1-1)
10.09.2004	FC Haka	HJK	0.54	0.26	0.20	1.97	0.72	2.61	0.80	0.62	0.23	0.15	1 (1-0)
12.09.2004	RoPS	FC KooTeePee	0.49	0.27	0.24	1.70	1.47	2.18	1.82	0.41	0.24	0.35	2 (1-3)
12.09.2004	Tampere United	FF Jaro	0.63	0.23	0.14	1.90	0.86	2.49	0.99	0.57	0.24	0.19	1 (2-0)
13.09.2004	AC Allianssi	TP-47	0.64	0.23	0.13	1.99	0.96	2.65	1.11	0.57	0.23	0.20	1 (3-2)
13.09.2004	TPS	FC Haka	0.24	0.27	0.49	0.77	1.03	0.86	1.21	0.25	0.34	0.41	X (0-0)
14.09.2004	HJK	FC Lahti	0.50	0.27	0.23	1.20	0.91	1.44	1.04	0.40	0.31	0.29	X (2-2)
15.09.2004	Tampere United	FC Hämeenlinna	0.72	0.18	0.10	2.22	0.90	3.04	1.03	0.62	0.21	0.17	1 (1-0)
18.09.2004	FC Haka	MyPa	0.56	0.26	0.18	1.61	0.69	2.03	0.77	0.55	0.27	0.18	1 (4-3)
18.09.2004	FC Hämeenlinna	HJK	0.22	0.26	0.52	1.13	1.41	1.34	1.73	0.29	0.28	0.43	X (2-2)
18.09.2004	FC Inter	Tampere United	0.36	0.28	0.36	1.25	0.98	1.50	1.14	0.40	0.30	0.30	1 (1-0)
18.09.2004	FF Jaro	FC Lahti	0.36	0.28	0.36	1.18	1.63	1.41	2.06	0.27	0.26	0.47	1 (2-1)
18.09.2004	FC Jazz	TPS	0.31	0.28	0.41	1.18	1.24	1.41	1.50	0.33	0.29	0.38	X (1-1)
18.09.2004	FC KooTeePee	AC Allianssi	0.33	0.28	0.39	1.28	1.31	1.55	1.60	0.34	0.28	0.38	2 (0-2)
22.09.2004	AC Allianssi	Tampere United	0.40	0.29	0.31	1.42	1.06	1.75	1.24	0.42	0.28	0.30	X (0-0)

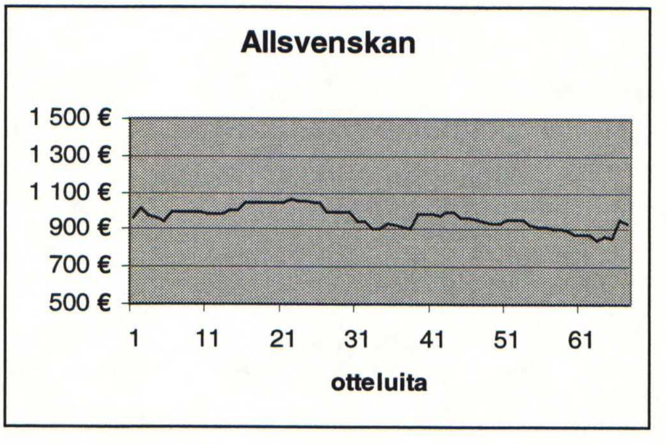
			MARKKINAT			BVDIP-malli								TULOS
			todennäköisyydet						maalien odotusarvot		todennäköisyydet			
Päivä	Kotijoukkue	Vierasjoukkue	1	X	2	λ_h	λ_a	λ_3	E[h]	E[a]	1	X	2	
05.07.2004	RoPS	Lahti	0.31	0.27	0.42	1.07	1.43	0.00	1.07	1.41	0.27	0.30	0.43	X (2-2)
05.07.2004	Jazz	Allianssi	0.27	0.27	0.46	1.45	1.06	0.00	1.43	1.06	0.44	0.30	0.26	X (1-1)
05.07.2004	Hämeenlinna	Inter	0.24	0.27	0.49	0.86	1.66	0.00	0.87	1.63	0.18	0.28	0.54	2 (3-4)
05.07.2004	HJK	TPS	0.56	0.25	0.19	1.53	1.01	0.00	1.51	1.01	0.47	0.29	0.24	1 (3-2)
05.07.2004	Haka	KooTeePee	0.60	0.24	0.16	2.02	0.67	0.00	1.97	0.68	0.66	0.23	0.11	1 (4-1)
07.07.2004	TP-47	Tampere_Utd	0.28	0.28	0.44	1.16	1.24	0.00	1.15	1.23	0.33	0.31	0.36	1 (1-0)
08.07.2004	AC Allianssi	FC Haka	0.37	0.28	0.34	0.97	1.46	0.00	0.97	1.44	0.24	0.30	0.46	X (0-0)
08.07.2004	FC Inter	FF Jaro	0.62	0.23	0.14	2.88	0.97	0.00	2.79	0.97	0.72	0.18	0.09	1 (5-1)
08.07.2004	FC Lahti	TPS	0.44	0.28	0.28	1.43	1.21	0.00	1.41	1.20	0.40	0.29	0.30	2 (1-2)
08.07.2004	RoPS	HJK	0.17	0.26	0.57	0.88	1.53	0.00	0.88	1.51	0.20	0.29	0.50	1 (2-0)
11.07.2004	FC KooTeePee	FC Jazz	0.49	0.27	0.24	1.49	1.08	0.00	1.47	1.08	0.45	0.29	0.26	X (1-1)
11.07.2004	TP-47	FC Hämeenlinna	0.54	0.26	0.20	2.49	1.14	0.00	2.42	1.14	0.63	0.21	0.15	1 (2-0)
11.07.2004	TPS	MyPa	0.42	0.28	0.30	1.78	1.04	0.00	1.75	1.04	0.52	0.27	0.21	1 (1-0)
14.07.2004	RoPS	FC Hämeenlinna	0.46	0.27	0.27	1.86	1.52	0.00	1.82	1.50	0.44	0.26	0.30	X (2-2)
15.07.2004	MyPa	FC KooTeePee	0.51	0.27	0.22	1.61	1.01	0.00	1.58	1.01	0.49	0.28	0.23	X (1-1)
15.07.2004	FF Jaro	Tampere United	0.27	0.28	0.45	1.07	1.94	0.00	1.06	1.90	0.19	0.25	0.55	2 (0-3)
15.07.2004	TP-47	FC Lahti	0.41	0.28	0.31	1.67	1.14	0.00	1.64	1.14	0.47	0.28	0.25	1 (3-1)
18.07.2004	FC Jazz	HJK	0.25	0.27	0.48	1.33	1.12	0.00	1.32	1.12	0.39	0.31	0.30	X (0-0)
18.07.2004	FC KooTeePee	FF Jaro	0.55	0.26	0.19	2.59	0.98	0.00	2.51	0.98	0.68	0.20	0.12	1 (2-0)
18.07.2004	TPS	TP-47	0.49	0.27	0.24	1.72	0.92	0.00	1.68	0.92	0.54	0.27	0.19	1 (3-2)
18.07.2004	FC Haka	RoPS	0.69	0.20	0.11	2.58	0.66	0.00	2.51	0.68	0.74	0.18	0.07	1 (1-0)
21.07.2004	Tampere United	FC Haka	0.35	0.29	0.36	1.08	1.05	0.00	1.08	1.05	0.34	0.33	0.33	2 (1-3)
21.07.2004	FC Hämeenlinna	MyPa	0.31	0.29	0.40	1.45	1.61	0.00	1.43	1.59	0.33	0.27	0.40	2 (1-2)
21.07.2004	FF Jaro	TP-47	0.38	0.29	0.33	1.17	1.46	0.00	1.16	1.44	0.29	0.29	0.42	1 (3-0)
22.07.2004	FC Lahti	FC Inter	0.36	0.28	0.36	0.99	1.21	0.00	0.99	1.20	0.28	0.32	0.39	1 (2-1)
22.07.2004	RoPS	AC Allianssi	0.27	0.28	0.45	1.27	1.62	0.00	1.26	1.59	0.29	0.28	0.43	1 (1-0)
24.07.2004	FF Jaro	FC Hämeenlinna	0.46	0.28	0.26	1.76	1.67	0.00	1.72	1.64	0.39	0.26	0.35	X (1-1)
24.07.2004	HJK	FC KooTeePee	0.62	0.24	0.14	1.92	1.05	0.00	1.88	1.04	0.55	0.26	0.19	1 (3-2)
25.07.2004	MyPa	FC Lahti	0.44	0.28	0.27	1.51	1.05	0.00	1.48	1.05	0.45	0.29	0.25	X (0-0)
25.07.2004	TP-47	FC Jazz	0.48	0.28	0.24	1.57	1.33	0.00	1.55	1.32	0.41	0.28	0.31	1 (3-2)
29.07.2004	Tampere_Utd	Jazz	0.57	0.25	0.18	1.77	0.81	0.00	1.74	0.82	0.57	0.27	0.16	1 (6-0)
29.07.2004	Inter	TP-47	0.59	0.24	0.17	2.26	1.01	0.00	2.20	1.01	0.62	0.23	0.15	1 (2-1)
29.07.2004	MyPa	Jaro	0.55	0.25	0.20	2.11	1.04	0.00	2.06	1.04	0.59	0.24	0.17	1 (2-1)
01.08.2004	FC Inter	MyPa	0.55	0.26	0.19	1.86	1.04	0.00	1.82	1.03	0.54	0.26	0.20	2 (0-1)
01.08.2004	AC Allianssi	TPS	0.49	0.27	0.24	1.43	1.31	0.00	1.41	1.30	0.38	0.29	0.33	X (0-0)
01.08.2004	FC Hämeenlinna	FC Jazz	0.37	0.28	0.35	1.52	1.57	0.00	1.50	1.54	0.35	0.27	0.37	X (2-2)
01.08.2004	RoPS	Tampere United	0.28	0.28	0.44	0.94	1.69	0.00	0.94	1.65	0.20	0.28	0.52	1 (2-0)
05.08.2004	FC Haka	FC Inter	0.51	0.27	0.22	1.70	0.74	0.00	1.66	0.76	0.57	0.27	0.15	1 (4-0)
05.08.2004	FF Jaro	AC Allianssi	0.29	0.28	0.43	1.42	1.88	0.00	1.40	1.84	0.28	0.26	0.46	2 (1-2)
05.08.2004	FC Lahti	Tampere United	0.33	0.29	0.38	1.18	1.18	0.00	1.17	1.17	0.34	0.31	0.34	1 (2-1)
05.08.2004	RoPS	TP-47	0.39	0.29	0.32	1.56	1.51	0.00	1.53	1.48	0.37	0.28	0.35	X (1-1)
08.08.2004	FC KooTeePee	FC Hämeenlinna	0.55	0.26	0.19	2.10	1.15	0.00	2.05	1.14	0.56	0.24	0.19	1 (1-0)
08.08.2004	TP-47	HJK	0.30	0.29	0.41	1.54	1.42	0.00	1.52	1.40	0.39	0.28	0.33	2 (1-2)
08.08.2004	TPS	RoPS	0.57	0.25	0.18	2.31	1.04	0.00	2.25	1.04	0.62	0.22	0.15	1 (1-0)
11.08.2004	Tampere United	TP-47	0.59	0.25	0.16	1.99	0.88	0.00	1.94	0.88	0.60	0.25	0.15	1 (3-2)
12.08.2004	HJK	FC Inter	0.51	0.27	0.22	1.82	1.15	0.00	1.78	1.14	0.51	0.27	0.23	2 (0-1)
12.08.2004	FC Lahti	FC Jazz	0.54	0.26	0.20	1.93	1.25	0.00	1.88	1.24	0.51	0.26	0.23	1 (4-0)
12.08.2004	MyPa	RoPS	0.60	0.25	0.16	2.14	1.06	0.00	2.09	1.06	0.59	0.24	0.17	2 (1-2)
12.08.2004	TPS	FC KooTeePee	0.53	0.27	0.21	1.88	1.02	0.00	1.84	1.02	0.55	0.26	0.19	1 (1-0)
20.08.2004	HJK	AC Allianssi	0.50	0.27	0.23	1.76	1.02	0.00	1.72	1.02	0.52	0.27	0.21	2 (0-1)
21.08.2004	TP-47	FC Haka	0.22	0.27	0.51	0.82	1.49	0.00	0.83	1.47	0.20	0.30	0.51	2 (1-4)
22.08.2004	FC Inter	FC KooTeePee	0.59	0.25	0.17	1.77	0.76	0.00	1.74	0.77	0.59	0.26	0.15	X (1-1)
22.08.2004	Tampere United	TPS	0.51	0.27	0.22	1.46	1.05	0.00	1.44	1.05	0.44	0.30	0.26	1 (3-1)
22.08.2004	FC Hämeenlinna	FC Lahti	0.24	0.27	0.49	1.19	1.48	0.00	1.18	1.46	0.29	0.29	0.42	2 (0-2)
26.08.2004	FC KooTeePee	Tampere United	0.30	0.28	0.42	1.47	1.35	0.00	1.45	1.33	0.38	0.29	0.33	2 (0-1)
26.08.2004	MyPa	TP-47	0.59	0.25	0.16	2.10	1.01	0.00	2.05	1.01	0.59	0.24	0.16	1 (2-0)
26.08.2004	TPS	FC Hämeenlinna	0.68	0.21	0.12	2.42	1.04	0.00	2.35	1.04	0.64	0.22	0.14	2 (2-3)
29.08.2004	AC Allianssi	FC Lahti	0.46	0.27	0.27	1.40	1.08	0.00	1.38	1.07	0.42	0.30	0.28	2 (1-2)
29.08.2004	FC Haka	FC Jazz	0.70	0.19	0.11	2.91	0.72	0.00	2.82	0.73	0.77	0.16	0.06	1 (1-0)
29.08.2004	FF Jaro	HJK	0.29	0.28	0.43	1.30	1.79	0.00	1.28	1.75	0.27	0.27	0.47	X (2-2)
29.08.2004	RoPS	FC Inter	0.31	0.28	0.41	1.24	1.51	0.00	1.23	1.48	0.30	0.29	0.41	1 (4-0)
29.08.2004	Tampere United	MyPa	0.46	0.27	0.27	1.32	1.10	0.00	1.30	1.10	0.40	0.31	0.30	X (1-1)
02.09.2004	HJK	MyPa	0.45	0.28	0.27	1.23	1.02	0.00	1.22	1.02	0.39	0.32	0.29	X (0-0)
02.09.2004	FC Lahti	FC Haka	0.26	0.28	0.46	0.91	1.55	0.00	0.92	1.53	0.21	0.29	0.50	2 (1-2)
02.09.2004	TP-47	FC KooTeePee	0.47	0.28	0.26	1.58	1.07	0.00	1.56	1.06	0.47	0.29	0.24	1 (1-0)
02.09.2004	FC Inter	FC Jazz	0.61	0.24	0.15	2.14	0.95	0.00	2.08	0.95	0.61	0.23	0.15	1 (3-0)
05.09.2004	AC Allianssi	FC KooTeePee	0.58	0.25	0.17	1.84	0.98	0.00	1.80	0.98	0.55	0.26	0.19	X (0-0)
09.09.2004	FC Jazz	RoPS	0.42	0.28	0.30	1.37	1.15	0.00	1.35	1.14	0.40	0.30	0.30	2 (0-1)
09.09.2004	FF Jaro	TPS	0.35	0.28	0.38	1.28	1.73	0.00	1.27	1.70	0.27	0.27	0.46	1 (1-0)
09.09.2004	FC Hämeenlinna	AC Allianssi	0.23	0.27	0.50	1.10	1.43	0.00	1.10	1.41	0.28	0.30	0.42	2 (2-3)
09.09.2004	MyPa	FC Inter	0.47	0.27	0.26	1.46	0.90	0.00	1.43	0.91	0.48	0.30	0.22	X (1-1)
10.09.2004	FC Haka	HJK	0.54	0.26	0.20	1.99	0.74	0.00	1.95	0.76	0.63	0.24	0.12	1 (3-2)
12.09.2004	RoPS	FC KooTeePee	0.49	0.27	0.24	1.72	1.53	0.00	1.68	1.50	0.40	0.27	0.33	2 (1-3)
12.09.2004	Tampere United	FF Jaro	0.63	0.23	0.14	1.95	0.87	0.00	1.91	0.88	0.60	0.25	0.15	1 (2-0)
13.09.2004	AC Allianssi	TP-47	0.64	0.23	0.13	2.05	1.00	0.00	2.00	1.00	0.59	0.24	0.17	1 (3-2)
13.09.2004	TPS	FC Haka	0.24	0.27	0.49	0.83	1.09	0.00	0.84	1.08	0.26	0.34	0.39	X (0-0)
14.09.2004	HJK	FC Lahti	0.50	0.27	0.23	1.27	0.94	0.00	1.26	0.94	0.42	0.32	0.26	X (2-2)
15.09.2004	Tampere United	FC Hämeenlinna	0.72	0.18	0.10	2.27	0.90	0.00	2.21	0.91	0.65	0.22	0.13	1 (1-0)
18.09.2004	FC Haka	MyPa	0.56	0.26	0.18	1.66	0.71	0.00	1.63	0.72	0.57	0.28	0.15	1 (4-3)
18.09.2004	FC Hämeenlinna	HJK	0.22	0.26	0.52	1.17	1.48	0.00	1.16	1.46	0.29	0.29	0.42	X (2-2)
18.09.2004	FC Inter	Tampere United	0.36	0.28	0.36	1.32	1.03	0.00	1.31	1.03	0.41	0.31	0.28	1 (1-0)
18.09.2004	FF Jaro	FC Lahti	0.36	0.28	0.36	1.21	1.68	0.00	1.20	1.65	0.26	0.28	0.46	1 (2-1)
18.09.2004	FC Jazz	TPS	0.31	0.28	0.41	1.25	1.32	0.00	1.24	1.30	0.34	0.30	0.36	X (1-1)
18.09.2004	FC KooTeePee	AC Allianssi	0.33	0.28	0.39	1.33	1.35	0.00	1.31	1.34	0.35	0.29	0.36	2 (0-2)
22.09.2004	AC Allianssi	Tampere United	0.40	0.29	0.31	1.46	1.11	0.00	1.44	1.11	0.43	0.30	0.27	X (0-0)

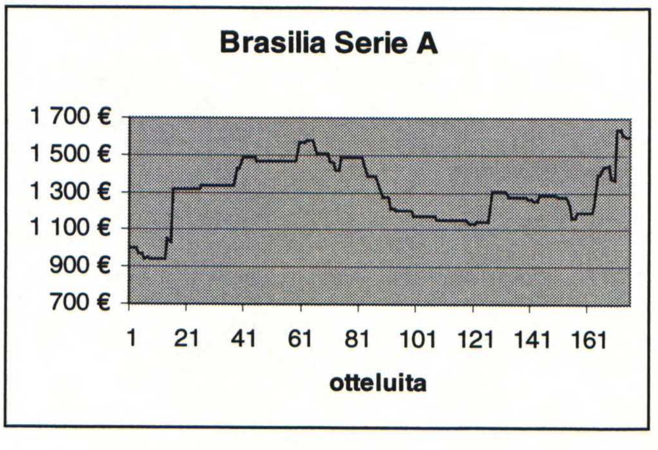
Liite B: Graafeja pelikassan kehityksestä eri sarjoissa

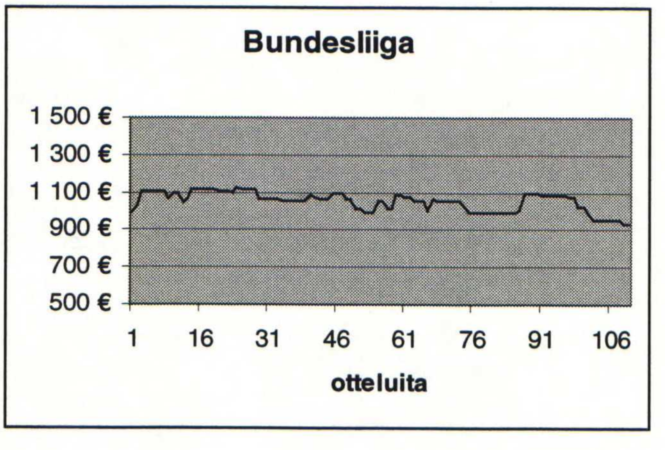


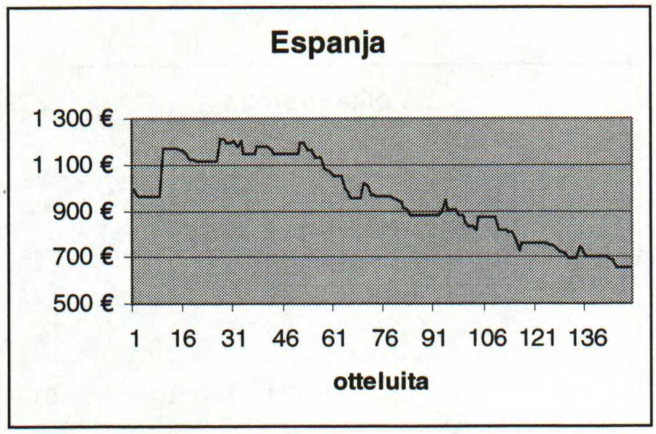


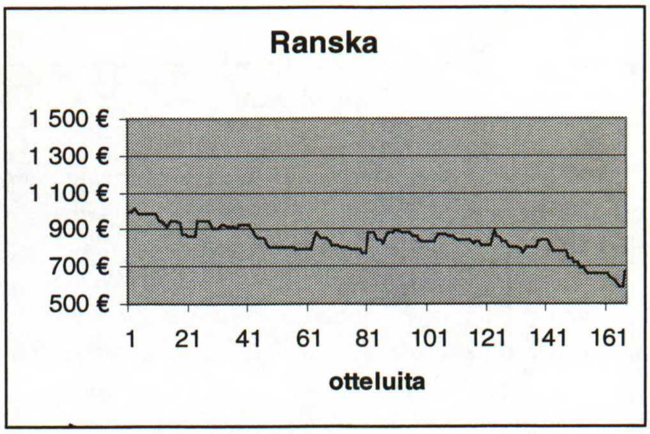


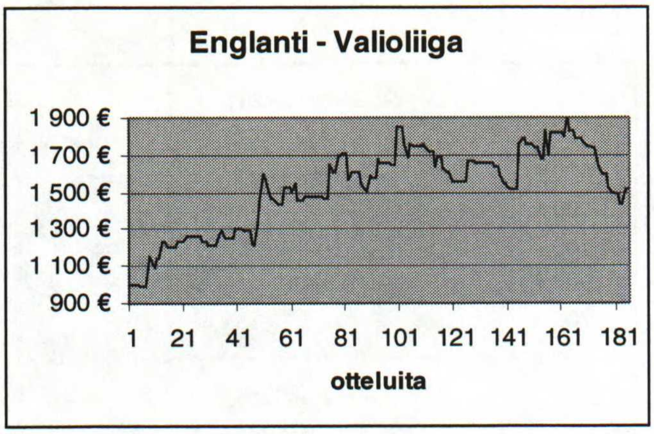


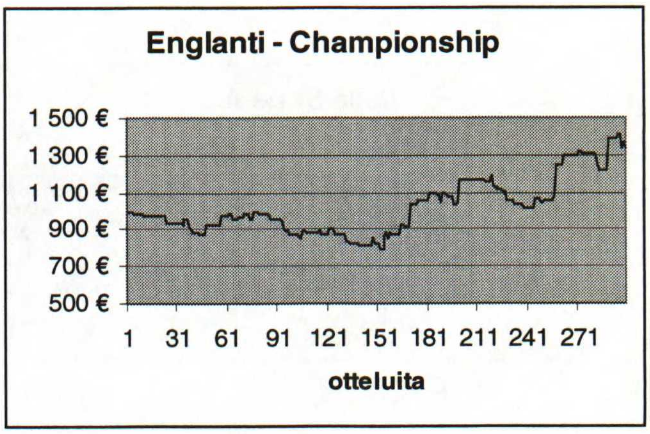


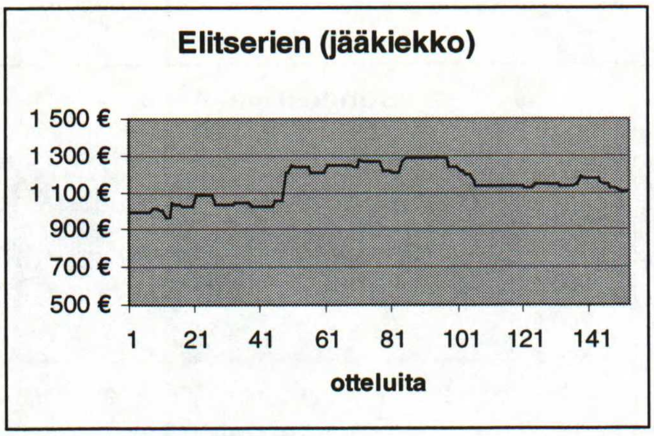


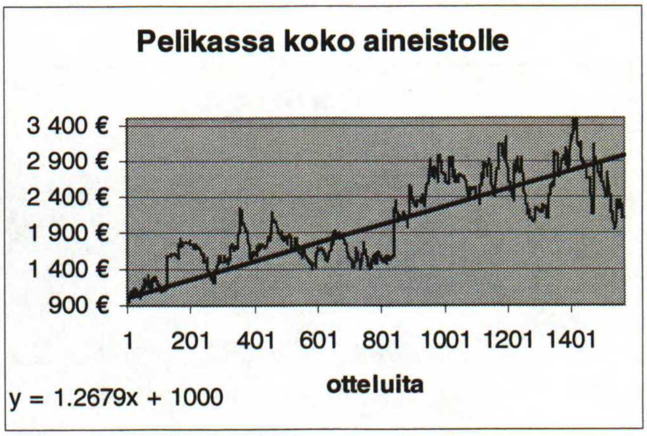

Espanja
1 300 €
1 100 €
900 €
700 €
500 €
1 16 31 46 61 76 91 106 121 136
otteluita


Ranska
1 500 €
1 300 €
1 100 €
900 €
700 €
500 €
1 21 41 61 81 101 121 141 161
otteluita


Englanti - Valioliiga
1 900 €
1 700 €
1 500 €
1 300 €
1 100 €
900 €
1 21 41 61 81 101 121 141 161 181
otteluita


Englanti - Championship
1 500 €
1 300 €
1 100 €
900 €
700 €
500 €
1 31 61 91 121 151 181 211 241 271
otteluita


Elitserien (jääkiekko)
1 500 €
1 300 €
1 100 €
900 €
700 €
500 €
1 21 41 61 81 101 121 141
otteluita


Pelikassa koko aineistolle
3 400 €
2 900 €
2 400 €
1 900 €
1 400 €
900 €
1 201 401 601 801 1001 1201 1401
otteluita
y = 1.2679x + 1000

Liite C: Rotatoidut faktorit ja muuttujien faktorilataukset [30]

MUUTTUJA	FAKTORI 1	2	3	4	5	6	7	8	9	10	11	12
koti_voittopros_kotona				0.896								
koti_tasuripros_kotona				-0.682						-0.682		
koti_tappiopros_kotona	-0.511									0.705		
koti_tehdyt_kotona				0.877								
koti_päästetyt_kotona	-0.558											
koti_pisteitä_kotona				0.834								
koti_voittopros_vieraissa	0.797											
koti_tasuripros_vieraissa							-0.937					
koti_tappiopros_vieraissa	-0.821											
koti_tehdyt_vieraissa												0.786
koti_päästetyt_vieraissa	-0.907											
koti_pisteitä_vieraissa	0.882											
koti_voittopros_yhteensä	0.716			0.579								
koti_tasuripros_yhteensä							-0.731					
koti_tappiopros_yhteensä	-0.844											
koti_tehdyt_yhteensä				0.625								0.541
koti_päästetyt_yhteensä	-0.919											
koti_pisteitä_yhteensä	0.808											
koti_voittopros_kuntopuntari					0.795							
koti_tasuripros_kuntopuntari							-0.559			-0.58		
koti_tappiopros_kuntopuntari					-0.734							
koti_tehdyt_kuntopuntari					0.705							
koti_päästetyt_kuntopuntari	-0.651				-0.543							
koti_pisteitä_kuntopuntari					0.84							
vieras_voittopros_kotona			0.882									
vieras_tasuripros_kotona			-0.573					-0.713				
vieras_tappiopros_kotona								0.537				
vieras_tehdyt_kotona			0.872									
vieras_päästetyt_kotona		-0.515										
vieras_pisteitä_kotona			0.859									
vieras_voittopros_vieraissa		0.824										
vieras_tasuripros_vieraissa									-0.958			
vieras_tappiopros_vieraissa		-0.799										
vieras_tehdyt_vieraissa											0.858	
vieras_päästetyt_vieraissa		-0.924										
vieras_pisteitä_vieraissa		0.898										
vieras_voittopros_yhteensä		0.69	0.62									
vieras_tasuripros_yhteensä								-0.633	-0.55			
vieras_tappiopros_yhteensä		-0.824										
vieras_tehdyt_yhteensä			0.667								0.538	
vieras_päästetyt_yhteensä		-0.893										
vieras_pisteitä_yhteensä		0.779	0.561									
vieras_voittopros_kuntopuntari						0.793						
vieras_tasuripros_kuntopuntari								-0.748				
vieras_tappiopros_kuntopuntari						-0.706						
vieras_tehdyt_kuntopuntari						0.697						
vieras_päästetyt_kuntopuntari		-0.675				-0.518						
vieras_pisteitä_kuntopuntari						0.828						

[30] Faktorilatauksista on karsittu pois ne, joiden itseisarvo on alle 0.5.